

## PRAISE FOR THE SECOND EDITION OF INTELLIGENT NETWORK VIDEO

---

“Fredrik Nilsson is a subject matter expert in the field of security technology. Having worked with him over the years, I continue to be impressed with his leadership and depth of knowledge, especially when it comes to the area of video surveillance systems. Fredrik believes in giving back to the security community and this particular work attests to this. In this increasingly volatile world, this work lends to the body of knowledge that all security professionals can benefit from. As an ‘end user,’ I found this work highly informative and valuable to me and my team. Read it!”

—Mike Howard, Chief Security Officer, Microsoft

“It’s become clear that harnessing advanced video technology is essential for providing an enhanced security environment. Video surveillance advances in technology, implementation, and application are all moving forward at a significant pace and allowing the security industry to provide more value than ever before. Fredrik’s knowledge, vision and insights are essential principles in understanding the video surveillance arena and I believe required reading to stay in the forefront of this topic for quality future-proof decision making. It’s definitely one of my go-to books on my reference shelf.”

—Brian J. Allen, Esq., CPP, CFE, CISSP, CISM, GVP, Chief Security Officer,  
Time Warner Cable, New York, NY

“Having been engaged in the security industry for more than 25 years, it’s very special to see a senior executive like Fredrik Nilsson commit time and energy to share his knowledge and experience as he does in his latest edition of *Intelligent Network Video*. Fredrik is broadly recognized and respected for his active support of industry-wide initiatives and as a thought leader in the area of industry education. His book is a comprehensive resource on the technology and applications of network video for surveillance systems. Building on his first edition, it addresses the product and technology advances in network video applications over the past several years and sets out what to expect next. Fredrik’s new edition is a timely update, and I recommend his book to everyone involved in the industry—for both industry veterans and those new to the industry.”

—Ken Boyda, Chairman of Razberi Technologies, Former CEO of Interlogix/GE Security

“When it comes to training apprentices, it is imperative that the best possible training material is used to educate the future of our industry! This book has the very best content of all of the books I have used and reviewed on the topic of video surveillance. It is very clear from the beginning of the book that the author understands how to convey complex topics to the reader and follows up with images and illustrations to reinforce the concepts. This book is by far the best resource available for people interested in learning about video surveillance!”

—Jim Simpson, Director of Telecommunications Curriculum, Electrical Training Alliance

“The global security industry is constantly evolving, influenced by rapid advancements in IP-based security technologies and external threats to business continuity such as cyber security threats. This makes it even more critical for industry professionals to not only stay current, but to position themselves a step ahead with technical knowledge that soars above commonly found industry information. As a former member of the Security Industry Association Board of Directors, Fredrik Nilsson has provided the industry with unique insight into the future direction of networked video and what it means for both security suppliers and practitioners. Through his writings, presentations, and commentary, Fredrik has dedicated considerable time and resources in support of all forms of industry education, from classroom instruction to online education and certification programs. This update to *Intelligent Network Video* is an example of his passion for industry education that contributes to a stronger industry for all stakeholders and the people and property they work to protect.”

—Don Erickson, Chief Executive Officer, Security Industry Association

“This book is a ‘must have’ for seasoned security professionals as well as freshly minted security system engineers as it tethers us to reason, facts and analysis, thus making it possible for us to make good technical decisions about the future.”

—**Pierre Racz, Founder and CEO, Genetec Inc.**

“In the highly competitive business of video surveillance, knowledge becomes power and it is a key differentiator. Whether you are beginning your career in the industry or are a veteran, whether you sell solutions or design and implement them, differentiate yourself by gaining extreme knowledge in video surveillance. Fredrik Nilsson’s **Intelligent Network Video, Second Edition** provides everything you need to know, and more, to shine above everyone else.”

—**Dan Mocer, Executive Chairman and Co-Founder, Convergent Technologies**



# Intelligent Network Video

Understanding Modern Video Surveillance Systems

S E C O N D   E D I T I O N



# Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

# Intelligent Network Video

Understanding Modern Video Surveillance Systems

S E C O N D   E D I T I O N

Fredrik Nilsson | Axis Communications



CRC Press

Taylor & Francis Group

Boca Raton London New York

---

CRC Press is an imprint of the  
Taylor & Francis Group, an **informa** business

CRC Press  
Taylor & Francis Group  
6000 Broken Sound Parkway NW, Suite 300  
Boca Raton, FL 33487-2742

© 2017 by Taylor & Francis Group, LLC  
CRC Press is an imprint of Taylor & Francis Group, an Informa business

No claim to original U.S. Government works

Printed on acid-free paper  
Version Date: 20160804

International Standard Book Number-13: 978-1-4665-5521-1 (Hardback)

This book contains information obtained from authentic and highly regarded sources. Reasonable efforts have been made to publish reliable data and information, but the author and publisher cannot assume responsibility for the validity of all materials or the consequences of their use. The authors and publishers have attempted to trace the copyright holders of all material reproduced in this publication and apologize to copyright holders if permission to publish in this form has not been obtained. If any copyright material has not been acknowledged please write and let us know so we may rectify in any future reprint.

Except as permitted under U.S. Copyright Law, no part of this book may be reprinted, reproduced, transmitted, or utilized in any form by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying, microfilming, and recording, or in any information storage or retrieval system, without written permission from the publishers.

For permission to photocopy or use material electronically from this work, please access [www.copyright.com](http://www.copyright.com) ([http://www.copyright.com/](http://www.copyright.com)) or contact the Copyright Clearance Center, Inc. (CCC), 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400. CCC is a not-for-profit organization that provides licenses and registration for a variety of users. For organizations that have been granted a photocopy license by the CCC, a separate system of payment has been arranged.

**Trademark Notice:** Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

---

#### Library of Congress Cataloging-in-Publication Data

---

Names: Nilsson, Fredrik, author.  
Title: Intelligent network video : understanding modern surveillance systems  
/ by Fredrik Nilsson.  
Description: New York : Routledge, 2017. | Includes bibliographical  
references and index.  
Identifiers: LCCN 2016022908 | ISBN 9781466555211 (hardback : alk. paper) |  
ISBN 9781466555235 (ebook)  
Subjects: LCSH: Video surveillance.  
Classification: LCC TK7882.E2 N55 2017 | DDC 621.389/28--dc23  
LC record available at <https://lccn.loc.gov/2016022908>

---

Visit the Taylor & Francis Web site at  
<http://www.taylorandfrancis.com>

and the CRC Press Web site at  
<http://www.crcpress.com>

# Contents

Acknowledgments	xvii
Introduction	xix
<b>1 Evolution of video surveillance systems</b>	<b>1</b>
1.1 VCR-based analog CCTV systems	2
1.2 DVR-based analog CCTV systems	2
1.3 Network DVR-based analog CCTV systems	3
1.4 Video encoder-based network video systems	4
1.4.1 Network video recorders and hybrid DVRs	4
1.5 Network camera-based network video systems	5
<b>2 Components of network video</b>	<b>7</b>
2.1 Where is network video used?	7
2.2 Network camera	9
2.2.1 Comparing a network camera and an analog camera	9
2.3 Video encoder	10
2.4 Network	12
2.5 Server and storage	12
2.6 Video management software	12
2.7 Intelligent video	14
<b>3 Network cameras</b>	<b>17</b>
3.1 Network camera components	17
3.2 Types of network cameras	18
3.2.1 Fixed cameras	18
3.2.2 Fixed dome cameras	19
3.2.3 PTZ cameras	20
3.2.4 Panoramic cameras	22
3.2.5 Covert cameras	22
3.2.6 Thermal cameras	23
3.3 PTZ cameras	24
3.3.1 Image stabilization	24
3.3.2 Presets and guard tours	24
3.3.3 Privacy masking	24
3.3.4 E-flip	25
3.3.5 Auto-flip	26
3.3.6 PTZ performance	27
3.3.7 Joystick control	27
3.4 Panoramic network cameras	28
3.4.1 Selecting the right viewing angle	28
3.4.2 Cameras with wide viewing angles	29
3.4.3 180° panoramic cameras	29
3.4.4 360° panoramic cameras	30
3.4.5 Multisensor panoramic cameras	32
3.4.6 Comparing and combining panoramic and PTZ cameras	32
3.5 Onboard cameras	34

3.6	Day-and-night network cameras	34
3.6.1	IR illuminators	34
3.6.2	Day-and-night applications	36
3.7	Megapixel network cameras	36
3.7.1	Benefits of megapixel	38
3.7.2	Megapixel applications	38
3.7.3	Drawbacks of megapixel	39
3.8	Best practices	40
4	Camera technologies	41
4.1	Light	41
4.1.1	Light characteristics	42
4.1.2	Illuminance	44
4.1.2.1	Definition of lux	44
4.1.2.2	Lux rating of network cameras	46
4.1.2.3	Lux rating of analog versus network cameras	47
4.1.3	Color temperature	49
4.1.4	Invisible light	49
4.2	Lenses	51
4.2.1	Lens types	51
4.2.1.1	IR-coated lenses	52
4.2.2	Lens mount standards	52
4.2.3	Field of view (focal length)	54
4.2.4	Matching lens and sensor	57
4.2.5	Aperture (iris diameter)	58
4.2.6	Types of iris control	60
4.2.7	F-number (f-stop)	61
4.2.8	Depth of field	62
4.2.9	Focusing	65
4.2.10	Lens quality	65
4.2.11	HDTV and megapixel lenses	67
4.3	Image sensors	67
4.3.1	Color filtering	67
4.3.2	CMOS and CCD technologies	69
4.3.2.1	CMOS technology	69
4.3.3	More about image sensors	70
4.3.4	HDTV and megapixel sensors	70
4.4	Image scanning techniques	71
4.4.1	Interlaced scanning	71
4.4.2	Deinterlacing techniques	71
4.4.3	Progressive scanning	73
4.5	Image processing	75
4.5.1	Exposure	75
4.5.2	Backlight compensation	76
4.5.3	WDR	76
4.5.3.1	Measuring dynamic range	78
4.5.3.2	Types of WDR	80
4.5.4	Bayer demosaicing	82
4.5.5	Noise	83
4.5.6	White balance	84
4.5.7	Sharpening and contrast	84
4.5.8	Aliasing	85

4.6	Resolution	86
4.6.1	NTSC and PAL resolutions	88
4.6.2	VGA resolutions	88
4.6.3	MPEG resolutions	89
4.6.4	Megapixel resolutions	90
4.6.5	HDTV resolutions	92
4.6.6	Ultra-HD resolutions	93
4.6.7	Aspect ratios	94
4.7	Best practices	95
5	Thermal cameras	97
5.1	How thermal imaging works	99
5.1.1	Electromagnetic spectrum	100
5.1.2	Near-infrared imaging	100
5.1.3	Using thermal radiation to create images	100
5.2	Components of a thermal camera	101
5.2.1	Sensors	102
5.2.1.1	Cooled sensors	102
5.2.1.2	Uncooled sensors	103
5.2.2	Sensor resolutions	104
5.2.3	Lenses for thermal cameras	104
5.2.3.1	Calculation of focal length	104
5.2.4	Thermal enclosures	105
5.3	Presentation of thermal images	106
5.3.1	Temperature alarm cameras	107
5.4	Determining detection range	108
5.4.1	Nomograph	109
5.4.2	Environmental considerations	109
5.4.2.1	Absorption	110
5.4.2.2	Scattering	110
5.5	Integrating thermal cameras with intelligent video	111
5.6	Export regulations for thermal technologies	112
5.7	Best practices	112
6	Video compression technologies	115
6.1	Basics of compression	115
6.1.1	Image and video compression	116
6.1.2	Lossless and lossy compression	117
6.1.3	Block transform	117
6.1.4	Prediction	117
6.1.5	Latency	117
6.1.6	Jitter	118
6.1.7	Compression ratio	118
6.2	Compression standards	118
6.2.1	ITU and ISO	118
6.2.2	History of compression formats	118
6.3	Compression formats	119
6.3.1	JPEG	119
6.3.2	Motion JPEG	121
6.3.3	JPEG 2000	121
6.3.4	Motion JPEG 2000	122
6.3.5	H.261 and H.263	122
6.3.6	MPEG-1	122
6.3.7	MPEG-2	123

6.3.8	MPEG-4	123
6.3.9	H.264	123
6.3.10	H.265	123
6.4	More on JPEG compression	124
6.5	More on MPEG compression	124
6.5.1	Frame types	124
6.5.2	Group of pictures	125
6.5.3	Constant, maximum, and variable bitrates	125
6.5.4	Profile@Level	127
6.5.5	Baseline and main profiles	127
6.5.6	Improving H.264 for surveillance needs	128
6.5.7	Licensing	128
6.5.8	Backward compatibility	129
6.6	Comparing standards	129
6.7	Best practices	130
<b>7</b>	<b>Audio technologies</b>	<b>131</b>
7.1	Audio modes	132
7.1.1	Simplex	133
7.1.2	Half duplex	133
7.1.3	Full duplex	133
7.2	Audio equipment	134
7.2.1	Audio input (microphones)	134
7.2.1.1	Condenser microphones	134
7.2.1.2	Electret condenser microphones	135
7.2.1.3	Dynamic microphones	136
7.2.1.4	Directional microphones	136
7.2.2	Audio output (speakers)	136
7.3	Acoustical adjustments	136
7.3.1	Volume and gain	136
7.3.2	Audio processing	137
7.3.3	Echo cancelation	137
7.3.4	Noise reduction	137
7.4	Audio detection alarm	137
7.5	Audio compression	138
7.5.1	Sampling rates	138
7.5.2	Bitrate	138
7.5.3	Software audio codecs	139
7.5.3.1	AAC-LC	139
7.5.3.2	G.711 PCM	139
7.5.3.3	G.726 ADPCM	139
7.5.3.4	G.722.2 or AMR-WB	140
7.6	Audio and video synchronization	140
7.7	The future of audio in network video	140
7.8	Other audio devices in network video systems	140
7.8.1	Network speakers	140
7.8.2	Network door station	141
7.8.3	SIP	142
7.9	Best practices	143
<b>8</b>	<b>Video encoders</b>	<b>145</b>
8.1	The components of a video encoder	145
8.2	Stand-alone video encoders	147



8.3	Rack-mounted video encoders	148
8.4	Video encoders with PTZ cameras	149
8.5	Video decoder	150
8.6	Best practices	151
<b>9</b>	<b>Wired networks</b>	<b>153</b>
9.1	Evolution of Ethernet	153
9.1.1	10 Mbit/s Ethernet	154
9.1.2	Fast Ethernet	154
9.1.3	Gigabit Ethernet	154
9.1.4	10 Gigabit Ethernet	155
9.1.5	Future of Ethernet	155
9.2	Network topologies	156
9.3	Network cabling	157
9.3.1	Twisted-pair cables and RJ45	157
9.3.2	Cable categories	158
9.3.3	Twisted-pair cable types	158
9.3.4	Fiber cable types	159
9.3.4.1	Fiber connectors	160
9.4	Basics of Ethernet	161
9.4.1	Media access control addresses	161
9.4.2	Frames	162
9.4.3	Half duplex and full duplex	162
9.5	Networking equipment	163
9.5.1	Hubs	163
9.5.2	Switches	163
9.5.3	Routers	164
9.5.4	Firewalls	165
9.5.5	Bridges	165
9.5.6	Internet connections	165
9.6	Power over Ethernet	165
9.6.1	802.3af and 802.3at standards	166
9.6.2	Midspans and splitters	167
9.7	Virtual local area networks	167
9.8	Best practices	168
<b>10</b>	<b>Wireless networks</b>	<b>171</b>
10.1	Basics of wireless networks	171
10.1.1	Wireless spectrum	172
10.1.2	Signal strength	174
10.1.3	Antennas	174
10.1.4	Radio wave propagation	175
10.2	Wireless network architectures	176
10.2.1	Point-to-point network	176
10.2.2	Point-to-multipoint network	176
10.2.3	Mesh network	177
10.3	802.11 WLAN standards	178
10.3.1	802.11b extension	178
10.3.2	802.11a extension	178
10.3.3	802.11g extension	178
10.3.4	802.11n extension	178
10.3.5	802.11ac extension	178
10.3.6	802.11s extension	178

10.4	Basics of 802.11 networks	179
10.4.1	Infrastructure network	179
10.4.2	802.11 frequencies	179
10.4.3	Channels	180
10.5	WLAN security	180
10.5.1	Wired equivalent privacy	180
10.5.2	Temporal key integrity protocol	180
10.5.3	Advanced encryption standard	180
10.5.4	Preshared key	180
10.5.5	802.1X	181
10.5.6	WiFi Protected Access®	181
10.6	Other wireless solutions	181
10.6.1	Bluetooth®	181
10.6.2	Universal mobile telecommunications system	182
10.6.3	Wireless interoperability for microwave access	182
10.7	Performance of wireless networks	182
10.8	Best practices	182
11	Networking technologies	185
11.1	OSI reference model	185
11.1.1	Layer 1: The physical layer	186
11.1.2	Layer 2: The data-link layer	186
11.1.3	Layer 3: The network layer	186
11.1.4	Layer 4: The transport layer	186
11.1.5	Layer 5: The session layer	187
11.1.6	Layer 6: The presentation layer	187
11.1.7	Layer 7: The application layer	187
11.2	TCP/IP reference model	187
11.2.1	Internet protocol	187
11.2.2	IPv4 addresses	188
11.2.3	Subnets	189
11.2.4	Network address translation	190
11.2.5	Services and port numbers	190
11.2.6	Port forwarding	191
11.2.7	IPv6	192
11.2.7.1	IPv6 addresses	192
11.3	Managing IP addresses	194
11.3.1	Setting IP addresses	194
11.3.2	Manual address allocation	194
11.3.2.1	Dynamic address allocation	194
11.3.3	Configuration-free networking	195
11.3.3.1	UPnP® and Zeroconf	196
11.3.3.2	Bonjour®	196
11.3.3.3	MAC and IP address resolution	196
11.3.3.4	Address resolution protocol	196
11.3.4	Domain name system	197
11.3.4.1	Dynamic DNS	198
11.4	Data transport	198
11.4.1	User datagram protocol	198
11.4.2	Transmission control protocol	198
11.5	Application layer protocols	199
11.5.1	Hypertext transfer protocol	199
11.5.2	File transfer protocol	199

11.5.3	Simple network management protocol	200
11.5.4	Simple mail transfer protocol	200
11.5.5	Real-time transport protocol	200
11.5.6	Session initiation protocol	200
11.6	Unicast, broadcast, and multicast	200
11.7	Quality of service	201
11.7.1	Definition	202
11.7.2	QoS in network video	202
11.8	Network security	203
11.8.1	Username and password authentication	203
11.8.2	IP filtering	203
11.8.3	802.1X	204
11.8.4	Virtual Private Network	205
11.8.5	Hypertext transfer protocol secure	206
11.9	Best practices for network security	206
11.9.1	Hardening guides	207
11.9.2	Best practices	207
12	<b>Servers and storage</b>	<b>209</b>
12.1	Servers	209
12.1.1	Hardware platforms	209
12.1.2	Operating systems	210
12.1.3	Video file systems	210
12.2	Hard disks	211
12.2.1	Small computer system interface	211
12.2.2	Advanced technology attachment and serial advanced technology attachment interfaces	211
12.2.3	Hard disk failure rates	212
12.2.4	Solid-state drives	212
12.3	Storage architecture	212
12.3.1	Edge storage	212
12.3.2	Single-server storage	214
12.3.3	Network-attached storage	214
12.3.4	Storage area network	214
12.3.5	Internet small computer system interface	215
12.4	Redundancy	215
12.4.1	Redundant array of independent disks systems	215
12.4.2	Data replication	216
12.4.3	Tape backup	216
12.4.4	Server clustering	217
12.4.5	Multiple servers	217
12.5	Best practices	217
13	<b>Video management</b>	<b>219</b>
13.1	Video management architectures	220
13.1.1	Server-based video management	220
13.1.1.1	PC server with VMS software	221
13.1.1.2	Network video recorder	222
13.1.2	Edge-based video management	223
13.1.3	Cloud-based video management	224
13.2	Other aspects of video management architecture	225
13.2.1	Open vs. vendor-specific software	225
13.2.2	Protocols and application programming interfaces	225
13.2.3	Apps for smartphones and tablets	225

13.2.4	Scalability of video management software	226
13.2.5	Licensing of video management software	226
13.3	System features	227
13.3.1	Recording	227
13.3.1.1	Video recording	227
13.3.1.2	Audio recording	228
13.3.1.3	Recording and storage	228
13.3.1.4	Search options	229
13.3.1.5	Exporting files	229
13.3.2	Viewing	230
13.3.2.1	Live viewing	230
13.3.2.2	Viewing of recordings	231
13.3.2.3	Multistreaming	231
13.3.2.4	Mapping functionality	232
13.3.3	Event management	233
13.3.3.1	Edge-based event handling	233
13.3.3.2	Responses	234
13.3.3.3	Video motion detection	235
13.3.3.4	I/O ports	237
13.3.3.5	Event log files	237
13.3.4	Administration and management features	238
13.3.4.1	Managing cameras	238
13.3.4.2	Time synchronization	238
13.3.4.3	Security	240
13.3.4.4	Audit log files	240
13.4	Integrated systems	241
13.4.1	Application programming interface	241
13.4.2	Point of sale	242
13.4.3	Physical access control	243
13.4.4	Building management	243
13.4.5	Industrial control systems	244
13.4.6	Radio-frequency identification	244
13.5	Best practices	245
14	Hosted video solutions	247
14.1	Principles of hosted video	247
14.2	Stakeholders of hosted video	249
14.2.1	Video hosting provider	250
14.2.2	Video service provider	250
14.2.3	Installer	250
14.2.4	Video service subscriber	250
14.3	Setting up hosted video	252
14.4	Characteristics of hosted video	253
14.4.1	Different needs, different services	255
14.4.1.1	Basic video surveillance	255
14.4.1.2	Alarm monitoring	256
14.4.1.3	Business intelligence	256
14.5	Data security	258
14.5.1	Audits, laws, and certifications	258
14.5.1.1	Standards for Attestation Engagements	258
14.5.1.2	ISO/IEC 27001 standard	259
14.5.1.3	Federal Information Security Management Act	259
14.5.1.4	European Union Agency for Network and Information Security	259

14.6	Integration with other systems	259
14.6.1	Integration with central station automation	259
14.6.2	Integration with physical access control	260
14.7	Best practices	261
15	Intelligent video	263
15.1	What is intelligent video?	263
15.2	Genesis of intelligent video	264
15.3	Why intelligent video?	265
15.3.1	Streamlining video surveillance operations	265
15.3.2	Managing stored video effectively	265
15.3.3	Improving business operations	266
15.4	Intelligent video architectures	267
15.4.1	Centralized systems	267
15.4.1.1	DVR-based installations	268
15.4.1.2	PC server-based installations	268
15.4.2	Distributed systems	268
15.4.2.1	Intelligence-at-the-edge installations	268
15.4.3	Integrating intelligent video applications	271
15.5	APIs and standards	272
15.5.1	Metadata	272
15.6	Best practices	272
16	Intelligent video applications	275
16.1	Categorizing video analytics	275
16.1.1	Categorizing video analytics by technology	275
16.1.1.1	Pixels, blobs, and objects	276
16.1.2	Categorizing video analytics by use	276
16.2	Analytics for security	277
16.2.1	Video motion detection	277
16.2.1.1	Evolution of VMD	277
16.2.1.2	Tuning of VMD parameters	277
16.2.2	Camera tampering detection	278
16.2.3	Object tracking	280
16.2.3.1	Crossline detection	280
16.2.3.2	Intrusion detection	280
16.2.3.3	Object left behind	281
16.2.3.4	Loitering detection	281
16.2.4	Fire and smoke detection	281
16.3	Analytics for business intelligence and operations	282
16.3.1	Object classification	283
16.3.2	Object and people counting	285
16.3.2.1	Customer traffic monitoring	286
16.3.2.2	Queue management	286
16.3.2.3	Tailgating	287
16.3.3	Dwell time and heat mapping	288
16.3.4	Traffic management	289
16.3.4.1	Incident detection	289
16.4	Hybrid analytics	290
16.4.1	Autotracking	291
16.4.2	Autotracking using PTZ	291
16.4.2.1	Gatekeeper	291
16.4.3	License plate recognition	292
16.4.4	Facial recognition	294

16.5	Video analytics and privacy	295
16.5.1	Sound identification	295
16.6	Realistic expectations on video analytics	296
16.7	Best practices	296
16.7.1	Video image quality	297
16.7.2	Efficient intelligent video algorithms	297
16.7.3	Computer processing power	297
16.7.4	Configuring and fine-tuning the system	298
17	<b>System design considerations</b>	<b>299</b>
17.1	Selecting a network camera	299
17.1.1	Types of camera	300
17.1.2	Image quality	300
17.1.3	Resolution	301
17.1.3.1	Determining the resolution needed	301
17.1.4	Compression	305
17.1.5	Networking functionality	305
17.1.6	Other functionalities	306
17.1.7	Vendor	307
17.2	Installing a network camera	308
17.2.1	Surveillance objective	308
17.2.2	Use plenty of light or add light if needed	308
17.2.2.1	Use IR light when white light is impracticable	309
17.2.3	Avoid direct sunlight and glare	310
17.2.4	Avoid backlight	310
17.2.5	Lens selection	312
17.3	Protecting a network camera	313
17.3.1	Camera enclosures in general	314
17.3.2	Transparent coverings	315
17.3.2.1	Overcoming the limitations of conventional domes	316
17.3.3	Positioning of fixed cameras	319
17.3.4	Environmental protection	319
17.3.5	Vandal and tampering protection	320
17.3.5.1	Goals of vandal protection	321
17.3.5.2	Mechanical design	321
17.3.5.3	Mounting	321
17.3.5.4	Camera placement	322
17.3.5.5	Intelligent video protecting cameras	323
17.3.6	Mounting types	324
17.3.6.1	Ceiling mounts	324
17.3.6.2	Wall mounts	324
17.3.6.3	Pole mounts	324
17.3.6.4	Parapet mounts	324
17.3.6.5	Special mounts	325
17.3.7	EMC	328
17.3.7.1	EMC standards	329
17.3.7.2	Emission	330
17.3.7.3	Immunity	330
17.3.7.4	Choosing between shielded and unshielded network cables	331
17.3.8	Safety of electrical equipment	331
17.3.9	Environmental ratings	332
17.3.9.1	IP ratings	332
17.3.9.2	NEMA ratings	334

	17.3.9.3	IK ratings	334
	17.3.9.4	IECEX and ATEX certifications	336
17.4		Storage and server considerations	337
	17.4.1	Small system: From 1 to 10 cameras	338
	17.4.2	Midsize system: From 10 to 100 cameras	338
	17.4.3	Large system: From 100 to 1000+ cameras	338
	17.4.4	Federated systems	338
	17.4.5	Provisioning the server	338
	17.4.6	Calculating storage	340
	17.4.6.1	Calculating storage needs	341
17.5		Provisioning network bandwidth	343
	17.5.1	Limiting the bandwidth	343
	17.5.2	Network and system latency	343
	17.5.3	Network cabling	344
	17.5.3.1	Tips for better network cabling	344
	17.5.3.2	Preparing the network cable	345
	17.5.3.3	Certifying the cable installation	345
17.6		Tools for designing systems	345
	17.6.1	Calculators and component selection tools	346
	17.6.2	Comprehensive system design tools	347
	17.6.3	Extensions for CAD software	348
	17.6.3.1	SketchUp®	349
	17.6.3.2	Revit®	349
	17.6.3.3	Visio®	349
17.7		Legal aspects	350
Author			353
Index			355



# Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>



# Acknowledgments

Writing a book, like any large project you take on in life, cannot be completed without tremendous support from colleagues, friends, and family.

I am fortunate enough to work for Axis Communications, a company that saw the value in the education this book provides. Axis not only gave me the time to work on this project but, more importantly, all the necessary internal resources to complete it. Special thanks go out to Shi-lin Chan, who acted as my shadow writer and editor for the first edition, and to Gunilla Burke, who helped me with this second edition. Also, thanks to Mark Listewnik of CRC Press, who (very) patiently waited for the manuscript. Many other people should be mentioned, but the list would become too long. Everyone who has helped me with this endeavor, from proofreading to writing whole chapters, is aware of your invaluable contributions and my sincere appreciation goes out to all of you.

I am also lucky to be part of an industry filled with outstanding individuals who entered the profession not only to make a living but also to help provide a safer and more secure life for all of us. Since moving to the United States 14 years ago, my knowledge of the industry and technology has grown tremendously, thanks to so many generous colleagues and customers willing to share their lifelong experiences with me.

And finally, this book would not exist without the understanding and support from my wife. Thanks for letting me spend so many weekends and nights on this project. I love you.



# Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

# Introduction

2016 marks the twentieth anniversary of the world's first IP network camera, the AXIS NetEye. What began as an experiment in streaming live images to web browsers came to change the video surveillance landscape. It has been eight years since *Intelligent Network Video: Understanding Modern Video Surveillance Systems* was first published by CRC Press. In that relatively short time-frame, technology has caused a profound shift in the way we live and work. Remember when smart-phones and social media were just emerging? Today we cannot imagine life without them. An internet that is still growing and the widespread adoption of cloud-based applications have enabled technology to permeate every facet of our lives—from online banking and shopping to telephony services, social interaction, and streamed entertainment.

Back in 2008, IP video was still an emerging technology. Skeptics wondered whether it was ready for primetime and if it would gain sufficient traction to eventually take over the whole video surveillance market. Video analytics was discussed as the biggest promise for the market, with estimates that it would become a billion-dollar market on its own within a few years. In the context of physical security, hardly anyone even talked about cloud services. IT security and physical security were mentioned in the context of “convergence,” but no one really considered the cyberthreat that IP video equipment on a corporate network could potentially present to a large organization.

Fast-forwarding to 2016: Today's IP cameras are light-years ahead of their analog predecessors. They are not only vastly better in terms of resolution, low-light capability, forensic information, and built-in intelligence, but they are also easier to install and have many more mounting options. Nowadays, we have tools and technology such as PoE (Power over Ethernet), outdoor readiness, and remote focus to greatly simplify installation. These advancements, as well as their immense scalability and open-standard design, are the reasons why enterprise-level projects today consider IP-based systems the only plausible option. Today, you can find systems with more than 10,000 network cameras integrated in a single surveillance system, and some systems even have more than 100,000 cameras. This level of scalability and seamless integration with other systems, such as access control, could only be possible using IP technology. And now thermal technology has entered the arena, adding a new dimension to IP surveillance. Given the potential of this technology, this edition has a new chapter on thermal network cameras.

One projection in the original edition that has yet to reach fruition is video intelligence or analytics. The current market is much smaller than what market researchers predicted back then. The sluggish growth to date stems from the general immaturity of the technology; the lack of relevant-sized, pure video analytics vendors; unclear patent situations; and the poor accuracy and value experienced by users. However, there are a few solid applications, such as people counting and license plate recognition, which are becoming mainstream.

It is interesting to note that, over the past eight years, video surveillance technology has changed and improved to a far greater extent than IT technology. Chapters 9 through 12 discuss some advancements in IT speeds and feeds, but they also show that the basic technology, architecture, and abilities are not that different from what they were eight years ago. There are, however, two new IT technologies that are influencing the physical security industry today: cybersecurity and cloud computing. With many high-profile data breaches over the last few years and cyberattacks between countries, cybersecurity is becoming a high priority for any corporation or government institution. This has meant that any physical security system running on the IT network by necessity has to adhere to stringent cybersecurity protocols. The other IT technology coming into play is cloud computing, also referred to as hosted video. While it has some great advantages, it is still an early

market, facing issues with bandwidth and storage scalability. Furthermore, business models for hosting and service providers are still being worked out. This edition also includes a new chapter on hosted video and its impact on the security industry (Chapter 14).

## WHAT TO EXPECT IN THE FUTURE

There is no need for a crystal ball to figure out that innovation will continue to reshape the surveillance landscape as we know it. During the last decade we have watched the security industry migrate from analog to almost fully IP based, a change in which IT departments that have become more accepting of video surveillance traffic on the network have played a role. Over the next 10 years, we will see changes even more dramatic, and the number of network camera installations is likely to escalate rapidly.

- *Technological improvements*

Inevitably, storage and bandwidth costs will drop so low that system designers will be free to build solutions that deliver live video at full frame rate to every user, and months or even years of archived recording. Improved video-compression technologies will help accelerate this trend further. Video quality will continue to improve, with HDTV becoming the de facto standard resolution. Most likely, it will first center on 1080p and then eventually on 4K resolution. With mobility and cloud offerings, access to video from anywhere will become commonplace.

- *Small- and midsize systems transitioning to IP*

For the reasons cited above, most enterprise systems installed today are IP based. Still, every year millions of analog cameras are sold globally—primarily to small- and midsize businesses. Industry analysis indicates that eventually all systems will become digital and IP based regardless of their size, but different drivers are in play in the small- and midsize markets. The more important factor in these markets is up-front cost. This is because most buyers only look at the cost of acquisition and do not evaluate the cost of operation and maintenance. However, as many new digital and IP-based systems are starting to address those concerns, the value conversation is changing. The small- and midsize markets challenge installers to provide additional value. In the future, as their technology and business practices mature, video hosting and recurring revenue streams are also expected to play an important role in these markets.

- *Consolidation of the market*

The market for manufacturers in the physical security equipment is very fragmented. Looking at the video space only, the top 15 manufacturers make up less than 50% of the market, which is quite unusual for most other markets. With the market becoming more and more IP based and the technology continuing to evolve at a rapid pace, manufacturers are pressed to invest greater resources in research and development (R&D) to keep up. This will most likely force the market to consolidate, which typically happens in two ways: via acquisitions or via “death.” In the latter case, companies go bankrupt or leave the video surveillance market to pursue other markets and a competitor steps in to take over their customer base.

For the end-customer, having fewer vendors to choose from is both good and bad. On a positive note, it might get easier to select the right partner. On the other hand, choices will be fewer and companies you partnered with a few years ago might no longer exist or no longer offer support for their older products and services.

- *Continued market growth*

The surveillance market has enjoyed healthy growth for many years. The underlying reason for that, unfortunately, is that we continue to face many threats both in our personal lives and in our business activities. This has necessitated that companies and governments alike spend time, money, and effort to provide security measures to mitigate those threats. According to research provided by the American Society for Industrial Security (ASIS), in the United States alone we spend an estimated \$1000 per capita a year on security, which adds up to \$300 billion,

or two percent, of our GDP. The annual total expenditure on security worldwide is, of course, significantly higher. How is that expenditure allocated? The vast majority of those funds are earmarked for human resources such as security guards, police, and other public protection services. Only a very small fraction is spent on electronic security equipment such as video surveillance cameras. This is likely to change though, just like many other markets that have become “automated” by technology.

To understand how technology has the potential to significantly elevate traditional security measures, it helps to compare its observational abilities to those of a human being. When surveillance cameras first entered the market, their job was simply to provide an extra pair of eyes, but we have now also harnessed their processing power, giving them both brains and memory. In a contest between man and machine, it quickly becomes evident that intelligent network video can amplify security well beyond what human intervention can achieve on its own. From visual acuity to attention span, from memory capacity to situational analysis, surveillance technology continues to help raise the bar for the security industry as a whole.

In recent years, public opinion has radically changed regarding the prevalence of surveillance technology in our lives. On April 30, 2013, TIME, CNN, and ORC International conducted a poll where 81% of respondents were in favor of surveillance cameras on streets and in public places. That is quite a significant shift in public opinion as compared to a mere 10 years ago, when 60% of respondents felt that way. This need for people to feel secure—whether at work, out shopping, on public transport, or in a critical infrastructure facility—has led to a burgeoning business at both ends of the spectrum: from mega-large surveillance systems to small-scale surveillance installations.

The current trend of cities deploying thousands of IP cameras within their borders and retailers installing hundreds of cameras in their stores is expected to continue and, in fact, accelerate. A growing number of surveillance cameras will be hosted and managed in the cloud. Finally, video intelligence will eventually deliver on its promise and drive IP video acceptance even further and faster than before. Additionally, expect other physical security markets such as access control, intercom, and audio to become fully scalable and integrated IP-based systems, adding further value to physical security and information management systems.



# Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>



**Fredrik Nilsson** is vice president of the Americas for Axis Communications, overseeing the company's operations in North and South America and serving on the global management team. In his 20-year career at Axis, he has undertaken various roles in both Sweden and the United States. Since assuming responsibility for the Americas in 2003, revenues have increased more than twentyfold. Nilsson has also been instrumental in leading the surveillance industry shift from analog closed circuit television to network video.

Nilsson has been an active participant on the SIA (Security Industry Association) Board of Directors and Executive Committee for many years. A trusted industry speaker, he has delivered lectures and keynotes at many conferences, including influential shows such as Securing New Ground, ASIS, ISC West, Expo Seguridad, and Interop. In 2016, Nilsson was selected to receive the prestigious 2016 George R. Lippert Memorial Award, presented to individuals for their long-term service to SIA and the security industry, the impact of their efforts on behalf of SIA and the industry, and their integrity, leadership, and diplomacy as demonstrated in industry dealings.

An oft-quoted source for top publications such as *The New York Times*, *USA Today*, *The Washington Post*, and *Svenska Dagbladet*, Nilsson has also appeared on television shows on CNN, CNBC, Fox News, and NECN. He has also written articles and been quoted in many security, IT, and business publications in the United States and globally.

Prior to Axis, Nilsson served as a product manager for ABB, a global leader in power and automation technologies. A graduate of the Lund Institute of Technology, he holds a master's degree in electrical engineering and postgraduate studies in economics.

## ABOUT THE BOOK

---

The revised edition of *Intelligent Network Video* is more comprehensive in every area than the first edition. There are also two new chapters on thermal cameras (Chapter 5) and hosted video (Chapter 11). The book takes the reader on a tour through the building blocks of intelligent network video—from imaging to network cameras and video encoders, through the IT technologies of network and storage and into video management, intelligent video, and system design. There are also a few chapters that are more technical in nature, included in this edition specifically for readers interested in delving deeper into the details. These chapters also include the word *technologies* in their title.

The purpose of the second edition is to trace the trajectory of video surveillance technology from its roots to its current state and into its potential future role in security and beyond. For the reader, it is an opportunity to explore what the latest technology has to offer, as well as gain some insight into the direction that surveillance will take us in the years ahead.



# Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>



# CHAPTER 1

## Evolution of video surveillance systems

Video surveillance, more commonly called closed-circuit television (CCTV), is an industry that emerged in the 1970s as a viable commercial technology to deliver safety and security. In the many years since then, the industry has experienced its share of technology changes. As in any other industry, end users' ever-increasing demands on products and solutions are driving this development, and evolving technologies are helping to support them. In the video surveillance market, the demands include:

- Better image quality
- Streamlined installation and maintenance
- More secure and reliable technology
- Better options for storing recorded video and greater capacity
- Lower cost and better return on investment
- Larger systems and better scalability
- Remote monitoring capabilities
- Integration with other systems
- Smarter video surveillance products—edge storage and built-in intelligence

To meet these requirements, video surveillance has experienced a number of technology shifts. The latest is the shift from analog CCTV surveillance to fully digital, network-based video surveillance systems.

Video surveillance systems started out as 100% analog systems and have gradually become digital, even though analog systems remain in use to some degree. Today's fully digital systems, which are using network cameras and PC servers for video recording, have come a long way from the early analog tube cameras, which were connected to VCRs that required hand-operated tape switching.

The existence of several semidigital solutions has led to some confusion in the video surveillance industry. Neither fully analog nor fully digital, they are partly digital systems that incorporate both digital and analog devices. Only systems in which video streams are continuously being transported over an IP network are truly digital and provide real scalability and flexibility. Still, some define semidigital systems with analog cameras connected to a digital video recorder (DVR) as digital systems, while others reserve the term for fully digital systems with network cameras and PC servers. Though all these systems contain digital components, there are some very important distinctions between them.

The following sections outline the evolution of video surveillance systems. Different system configurations are explained, from fully analog to fully digital, along with the benefits of each

configuration. Section 1.1 describes analog systems. Sections 1.2 and 1.3 describe semidigital video systems. Sections 1.4 and 1.5 describe true network video systems. However, only the system outlined in Section 1.5 is fully digital.

## 1.1 VCR-BASED ANALOG CCTV SYSTEMS

The traditional analog CCTV system comprised analog cameras that were connected to a VCR for recording video (Figure 1.1). The system was completely analog and the VCR used the same type of cassettes as those sold for a home VCR. Each camera was connected by its own coax cable to the VCR. The video was not compressed, and a tape's recording capacity at full frame rate was limited to a maximum of 8 hours.

Eventually, a so-called time-lapse mode was incorporated into VCRs to increase tape capacity. Timelapse enabled recording of every second, 4th, 8th, 16th, or 30th second image or frame. In analog systems that used time-lapse recording, these were the only recording frame rates possible, and that was how the video surveillance industry came up with specifications such as 30, 15, 7.5, 3.75, and 1.875 fps (frames per second). When several cameras were used, quads became another important system component. A quad simply took inputs from four cameras and created one video signal output to show four different images on one screen, hence the name quad. This invention made the system slightly more scalable but at the expense of lower resolution.

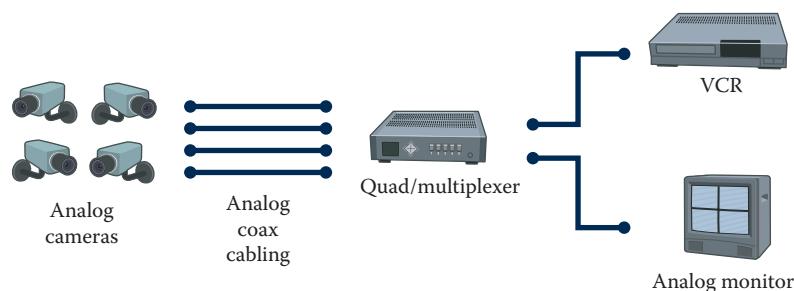
In larger systems, multiplexers became commonplace. A multiplexer combined the video signals from several cameras into a multiplexed video signal. This made it possible to record material from a larger number of cameras, often 16 on one device. The multiplexer also made it possible to map selected cameras to specific viewing monitors in a control room. But all equipment and all signals were still analog. For operators to observe the video, analog monitors had to be connected to a VCR, quad, or multiplexer.

Although analog systems functioned well, the drawbacks included limitations in scalability and the need to maintain VCRs and manually change tapes. In addition, the quality of the recordings deteriorated over time. For a long time, cameras could deliver only black-and-white images. Today, most analog cameras support color video.

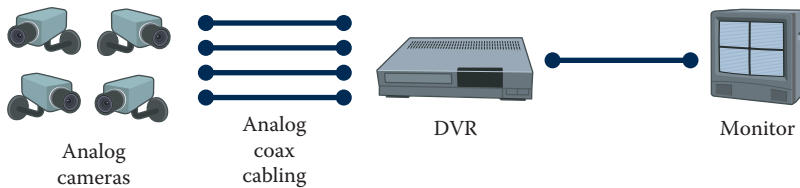
## 1.2 DVR-BASED ANALOG CCTV SYSTEMS

By the mid-1990s, the video surveillance industry saw its first digital revolution with the introduction of the DVR. In a DVR, the videotape is replaced with hard drives for the video recording, which requires the video to be digitized and compressed in order to store as many days' worth of video as possible (Figure 1.2).

Hard disk space was scarce in early DVRs, so either the recording duration had to be shorter or a lower frame rate had to be used. Limitations in hard disk space prompted many manufacturers to



**Figure 1.1** A classical analog video surveillance system.



**Figure 1.2** A surveillance system with analog cameras connected to a digital video recorder, including quad or multiplexer functionality and providing digital recording.

develop proprietary compression algorithms. Though these algorithms worked well, end users were tied to one manufacturer's tools when it came to playing back the video. Over the years, the cost of hard disk space decreased dramatically, and standard compression algorithms became available and widely accepted. To the benefit of end users, most manufacturers gave up their proprietary compression formats in favor of standards such as JPEG, MPEG-4, and H.264.

Most DVRs had multiple video inputs (typically 4, 16, or 32), which meant DVRs also included the functionality of the quad or multiplexer. DVRs replaced the multiplexer and the VCR, thereby reducing the number of components in CCTV systems.

The introduction of the DVR system delivered the following major advantages:

- No tapes or tape changes
- Consistent recording quality
- Ability to search quickly through recorded video

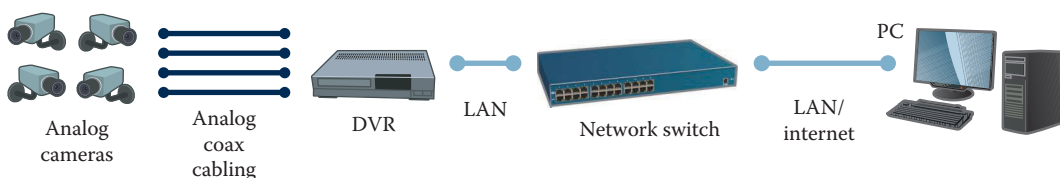
Early DVRs used analog monitors such as TV sets for showing video. However, because the DVR made digital video available, it became possible to network and transmit the digital video remotely through a phone modem connected to a serial port on the DVR. Later, the phone modem was built into the DVR itself. Though the ability to monitor video remotely on PCs was a great benefit, the actual functionality was not very useful because the bandwidth available with phone modems was too low—often in the 10–50 kbps (kilobits per second) range. That meant very low frame rates, low resolution, or highly compressed video, which made the video useless.

### 1.3 NETWORK DVR-BASED ANALOG CCTV SYSTEMS

DVRs were eventually equipped with an Ethernet port for network connectivity. This introduced network DVRs to the market and enabled remote video of live or recorded video using PCs (Figure 1.3). Some of today's systems require a special Windows® client to monitor the video, whereas others use standard web browsers that make remote monitoring more flexible.

The network DVR system provides the following advantages:

- Remote monitoring of video by PC
- Remote operation of the system



**Figure 1.3** An example of how analog cameras can be networked using a network digital video recorder for remote monitoring of live and recorded video.

Although DVRs provided great improvements over VCRs, they also had some inherent disadvantages. The DVR was burdened with many tasks, such as digitizing video from all cameras, compressing video, recording, and networking. In addition, it was a blackbox solution with proprietary hardware preloaded with software that often forced the end user to source spare parts from a single manufacturer, making maintenance and upgrading expensive. Virus protection was also difficult to implement. Though the DVR was often a Windows-based machine, the proprietary interface did not allow for virus protection. Moreover, the DVR could only offer limited scalability. Most DVRs offered 16 or 32 inputs that made it difficult to, in a cost-effective way, build systems that were not based on multiples of 16. For example, if the system required 10 or 40 cameras, the customer would be forced to buy a 16-channel DVR or a 48-channel DVR combination.

## 1.4 VIDEO ENCODER-BASED NETWORK VIDEO SYSTEMS

The first step into a network video system based on an open platform came with the introduction of the video encoder. Encoders are typically used in existing installations where analog cameras installed 2–4 years ago still function well, but where the DVR's shorter life span requires its replacement with a more flexible and future-proof recording solution.

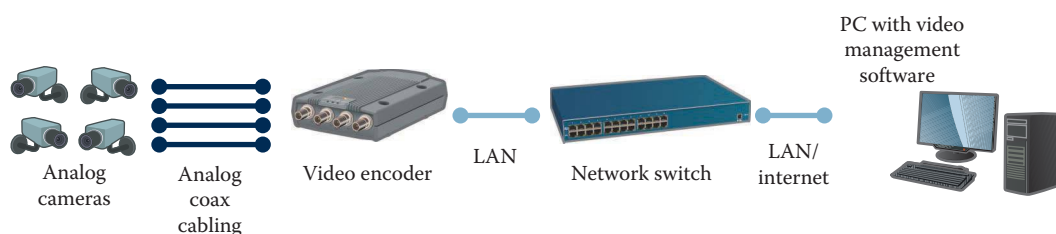
A video encoder connects to analog cameras and digitizes and compresses the video. It then sends the video over an IP network through a network switch to a PC server, which runs video management software for monitoring and recording (Figure 1.4). This is a true network video system because the video is sent continuously over an IP network. In essence, the tasks previously performed by the DVR are now divided up: the video encoder handles digitization and compression, and the PC server takes care of recording. One of the benefits of this solution is better scalability.

A video encoder-based network video system has the following advantages:

- *Nonproprietary hardware:* Standard network and PC server hardware can be used for video recording and management.
- *Scalability:* Cameras can be added one at a time because encoders are available in 1-, 4-, 16-, and up to 84-channel versions.
- *Off-site recording:* Record video remotely and at central locations.
- *Future-proof solution:* Systems are expanded easily by incorporating network cameras.

### 1.4.1 Network video recorders and hybrid DVRs

Alternatives to the open platform, which is based on a PC with installed video management software, are different types of network video recorders (NVRs) (Figure 1.5) and hybrid DVRs. An NVR or hybrid DVR is a proprietary hardware box with preinstalled video management software for managing video from video encoders or network cameras. The NVR handles only network video inputs, whereas the hybrid DVR can handle both network video and analog video inputs. Because recording and video management are available in a single unit, very much like the DVR, the benefit of using an NVR or hybrid DVR is the ease of installation. However, while they are



**Figure 1.4** A true network video system. Video is continuously transported over an IP network. It uses a video encoder as the cornerstone to migrate the analog security system into an open IP-based video solution.



**Figure 1.5** An example of a network video recorder with preinstalled video management software. Its installation is easy, but it lacks the flexibility of an open-platform system based on a standard PC server. (Image courtesy of Genetec™, Montreal, Quebec, Canada.)

often easy to install, NVRs and hybrid DVRs are often more difficult to maintain on a corporate IT network because they use proprietary platforms.

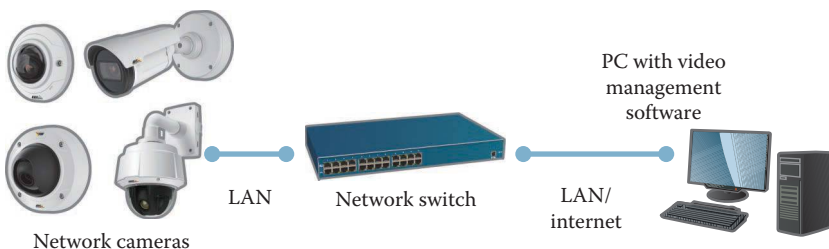
## 1.5 NETWORK CAMERA-BASED NETWORK VIDEO SYSTEMS

As its name indicates, a network camera—commonly called an IP camera—is a camera with an IP network connection. In a network camera-based video system, video is transported over an IP network through network switches and is recorded to a PC server with video management software installed (Figure 1.6). This represents a fully digital network video system.

One of the greatest benefits of a network camera is that when images are captured, they are digitized only one time (inside the camera) and then remain digitized throughout the system. As a result, the image quality is consistently high. This is not the case with analog cameras. Though most analog cameras today are called digital, they have an analog output, which can lead to some confusion. Analog cameras do digitize captured images to provide image-enhancing functions. However, these images are then converted back to analog video. It is important to know that with every conversion from analog to digital, or from digital to analog, there is some loss of video quality. Analog signals also degrade when transported over long cables and over time if stored on tape. Ideally, therefore, video should be digitized once and remain digital throughout the system.

Another advantage of IP-based networks is that one can use the network for more than just transporting video. Besides allowing several network cameras to share the same physical cable, IP networks can carry:

- Power to network cameras
- Information to and from the cameras' output and input contacts
- Pan, tilt, and zoom (PTZ) commands
- Two-way audio



**Figure 1.6** A true network video system where network camera video streams are continuously transported over an IP network. This system takes full advantage of digital technology and provides consistent image quality from the cameras to the viewer at any location.

In addition, an IP network enables network cameras to be configured remotely and allows video and other data sent over the network to reach virtually any location without reductions in quality. To sum it up, IP networks provide an extremely flexible and cost-effective medium for all communications within a network video surveillance system. Network video is scalable; it offers opportunities to build any size of video surveillance system, from a single camera to thousands of cameras.

A network camera-based network video system offers many advantages:

- Access to high-resolution (HDTV and megapixel) cameras.
- Consistent image quality, regardless of distances in the network.
- The same cabling can be used for video, power, PTZ, audio, and digital input and output, and audio.
- Access to wireless functionality.
- Remote access to camera settings and focus adjustments over IP.
- Access to edge intelligence and edge storage (i.e., built into the camera).
- Full flexibility and scalability.

Though an analog camera connected to a video encoder can be compared to a network camera, a network camera can offer many more value-adding functionalities.

The network camera is a key driver in the network video revolution. Network cameras have more than caught up with analog camera technology. Not only do they greatly surpass analog video quality in many aspects, such as resolution and light sensitivity, but they also offer the advantages of built-in intelligence and storage.

## CHAPTER 2

# Components of network video

To understand the scope and potential of an integrated, fully digitized system, let us first examine the core components of a network video system: the network camera, the video encoder, the network, the server and storage, and the video management software.

One of the main benefits of network video is that its network infrastructure platform consists of commercial off-the-shelf equipment. The network hardware and software, such as cables, servers, storage devices, and operating systems, is standard IT equipment. The other three components—the network camera, the video encoder, and the video management software—are unique to video surveillance and make up the cornerstones of network video solutions (see Figure 2.1). Another component is intelligent video, or video analytics, which can be built into or installed on the network camera or video encoder or be part of the video management software.

### 2.1 WHERE IS NETWORK VIDEO USED?

---

Network video, also commonly known as IP-based video surveillance or IP surveillance, is a system that gives users the ability to monitor and record video and audio over an IP network (local area network [LAN], wide area network [WAN], or the internet).

Unlike analog video systems, network video uses standard IP-based networks, as opposed to dedicated point-to-point cabling, as the backbone for transporting video and audio. In a network video system, digitized video and audio streams are sent over wired or wireless IP networks, enabling video monitoring and recording from anywhere on the network.

Network video can be used in an almost unlimited number of applications. However, because network video's advanced functionalities are highly suited for security surveillance, most applications belong to that category. The high-quality video, scalability, and built-in intelligence of network video enhance security personnel's ability to protect people, property, and assets.

Network video can also have many other purposes, such as the following:

- Remote monitoring of equipment, people, and places—locally and remotely. Application examples include traffic and production line monitoring.
- Business intelligence, especially in retail environments. Application examples include people counting, heat mapping, and monitoring multiple retail locations.



**Figure 2.1** Typical network video system components, including network cameras, video encoders, and video management software. The other components including the network, storage, and servers are all standard IT equipment.

- Broadcasting video, either over the internet or tying the video into broadcast production systems. Because the quality video of today's network cameras is HDTV compliant, it provides quality similar to that of broadcast camera equipment.
- Healthcare. Application examples include patient monitoring, assisting researchers in sleep-study labs, and providing a remote eye for a specialist doctor in an emergency care unit.

Like most new technologies, network video was first deployed in a few key market segments. The primary initial driver for network video was scalability, and therefore, most installations included hundreds or thousands of cameras in segments such as education, healthcare, and government. Today, network video is attractive for all market segments. And the interest is increasing even for the very small systems that historically have been addressed by analog systems.

The market segments where network video systems are being successfully installed include the following:

- *Education:* Secure and remote monitoring of school playgrounds, hallways, and classrooms in all levels of education, from K–12 through colleges and universities.
- *Transportation:* Surveillance of railway stations and tracks, parking lots and garages, highways, seaports, and airports, as well as mobile transportation environments such as buses, trains, and cruise ships.
- *Banking:* Traditional surveillance in bank branches, headquarters, and ATM locations.
- *Government:* Surveillance to ensure safe and secure government facilities such as state facilities, air force bases, courthouses, and correctional facilities.
- *City surveillance:* Surveillance to ensure safe environments in city centers for citizens and tourists and to help police do their job in a safer and more efficient way.
- *Retail:* Surveillance that makes store management easier and more efficient, including business intelligence applications such as people counting and integration with other systems such as point of sales. The retail market can be divided further, with large stores with over 50 cameras per store on one end and convenience stores that often have less than 8 cameras per store on the other.
- *Healthcare:* Surveillance that provides security for both staff and patients inside hospitals, on hospital campuses, and in parking garages. Network video is also increasingly finding its way



into applications other than traditional security within healthcare, such as drug diversion prevention and monitoring of patients with diminished mental or physical capacity.

- *Commercial:* Surveillance that provides safety for staff in midsize to large companies with thousands of employees and multiple locations, in office environments, building access points, warehouses, surrounding campus areas, and parking lots and garages.
- *Industrial:* Monitoring of the manufacturing processes, logistics systems, warehouses, and inventory control systems.

## 2.2 NETWORK CAMERA

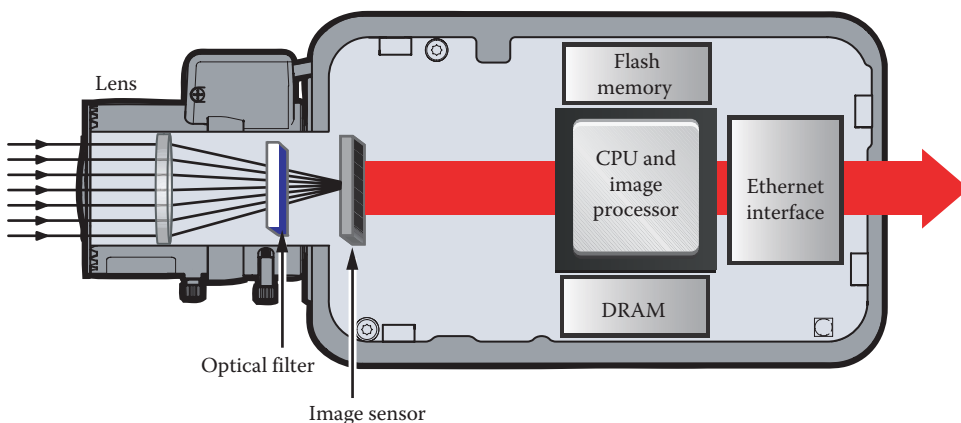
A network camera, often also called an IP camera, can be described as a camera and computer combined into one unit. It captures and sends live images directly over an IP network, enabling authorized users to locally or remotely view, store, and manage video over a standard IP-based network infrastructure (see Figures 2.2 and 2.3).

A network camera has its own IP address. It is connected to a network and has a built-in web server, FTP server, email client, intelligence, storage, programmability, and much more. A network camera operates as an independent server on a network and can be placed wherever there is an IP network connection. A web camera, or webcam, is something totally different. It is either a separate camera that only operates when it is connected to a computer through a USB port or integrated into the computer screen, smartphone, or tablet. To use a webcam, you need a software that can control it. Today, most people are familiar with webcams as they use them for video chatting.

### 2.2.1 Comparing a network camera and an analog camera

In recent years, network camera technology has not only caught up but widely surpasses all performance aspects of analog cameras. Network cameras also offer a number of advanced features that cannot be found in analog counterparts, such as high image quality, high resolutions, onboard storage, audio, and built-in intelligence.

Whereas an analog camera is a one-directional signal carrier that ends at the recording device, a network camera is fully bidirectional, allowing information to be sent and received. It can be an integrated part of a larger, scalable system, where it communicates with several software applications simultaneously to perform various tasks, such as detecting motion or sending different streams of video.



**Figure 2.2** Inside a network camera. The lens exposes the image onto an image sensor that is connected to a processor, often an application-specific integrated circuit or a digital signal processor. The camera has a flash memory and a dynamic random access memory and connects to the network through a network interface.



**Figure 2.3** Typical network camera with some important features pointed out. Front view (a). Back view (b).

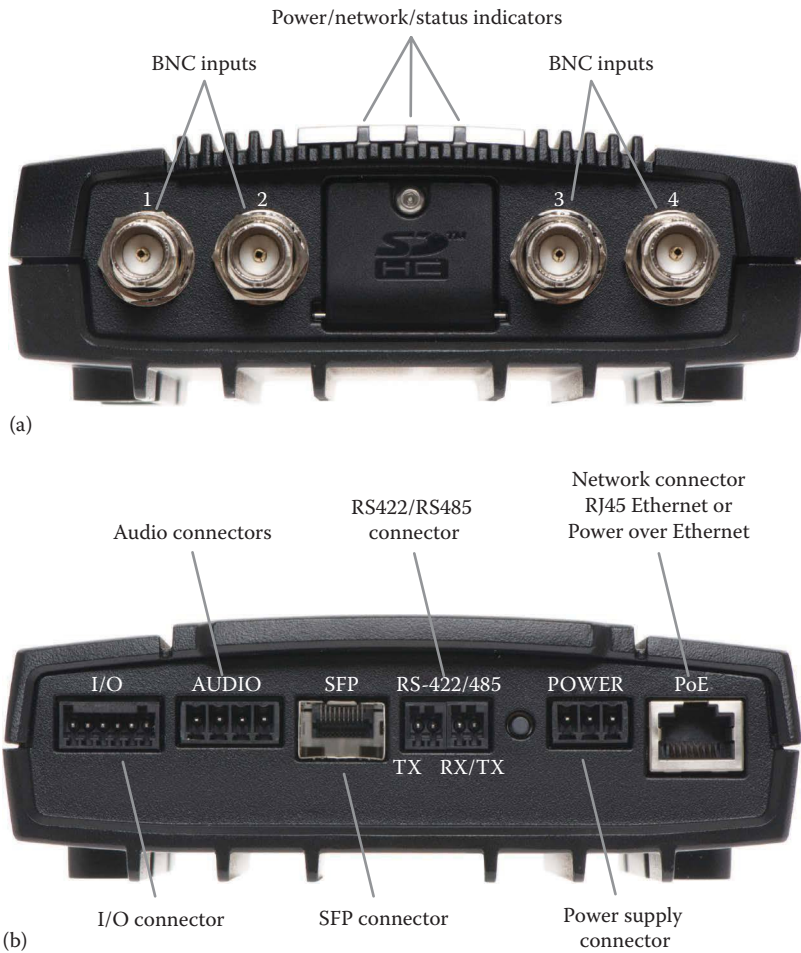
Read more about network cameras in Chapters 3 and 4. For more detailed information about the technology inside network cameras, including such topics as compression and audio, see Chapters 6 and 7.

## 2.3 VIDEO ENCODER

A video encoder makes it possible to integrate an analog camera into a network video system without having to discard existing analog equipment. It brings new functionalities to analog equipment and eliminates the need for dedicated equipment such as coaxial cabling, analog monitors, and DVRs—the latter becoming unnecessary as video recording can be done using standard PC servers.

A video encoder typically has 1, 4, or 16 ports for analog cameras to plug into, as well as an Ethernet port for connection to a network (Figure 2.4). Like network cameras, a video encoder contains a built-in web server, a compression chip, and an operating system so that incoming analog feeds can be converted into digital video and then sent to be recorded over a computer network for easier accessibility and viewing.

In addition to the video input, a video encoder can support other functionalities, including audio, alarm activation through digital inputs and outputs, and control of pan, tilt, and zoom (PTZ) mechanisms through serial ports or a coaxial cable. A video encoder also can be connected to a wide variety of specialized cameras, such as microscope cameras.



**Figure 2.4** Typical video encoder with some important features pointed out. Front view (a). Back view (b).

For large installations with existing analog cameras, rack-mounted versions of video encoders are available. Typically, one video encoder chassis offers 16 ports, but sometimes it offers as many as 84 ports (Figure 2.5).

With millions of analog cameras in operation, video encoders will continue to play an important role in video surveillance systems. Chapter 8 gives more information about video encoders.



**Figure 2.5** For large installations with coax cabling already installed, a video encoder chassis is a popular solution.

## 2.4 NETWORK

---

One of the most obvious benefits of network video is the ability to use standard wired or wireless IP-based networks. Not only are IP-based networks cost efficient to deploy, but in many cases they already exist and can be used also to carry power to a network device such as a network camera or video encoder. IP networks come in many different designs and sizes, from small wireless LANs in homes to corporate networks, citywide wireless deployments, and the internet. IP-based networks are used by almost everyone on a daily basis, at home or at work. Today, just like electrical wiring and plumbing, wiring for IP networks is included already at the planning stage of most buildings.

The growing use of IP networks means that companies are spending enormous amounts of money on research and development to constantly improve the functionality, speed, and security of such networks. Whereas bandwidth was a bottleneck some years ago, today's gigabit networks provide more than enough bandwidth for content-rich applications such as network video. Even WANs and mobile cellular networks are beginning to have sufficient bandwidth for video distribution at reasonable costs.

Because everyone relies on access to a network in their daily lives, network security is becoming an increasingly critical issue. It is especially so when the network carries sensitive data, such as bank transactions, government documents, and now also video surveillance streams. Therefore, it is important to understand network security technologies such as 802.1X, HTTPS, and firewalls.

Read more about wired and wireless networks in Chapters 9 and 10. Chapter 11 gives more details about network technologies and includes topics such as network security.

## 2.5 SERVER AND STORAGE

---

The server and storage components used in network video are based on standard IT equipment. This means that a network video system can benefit from rapid innovation in processor and storage technologies. The IT market is a huge market, 100 times bigger than the video surveillance market, with significant investments in research and development that result in ever-increasing server performance and storage capacity.

Most servers that run video management applications are PC servers with a Windows® operating system, and sometimes a Linux® or Unix™ operating system. Depending on what performance is required, a single or multiple processor server can be used. This makes it possible to manage video from hundreds of cameras per server. Enterprise-class servers usually deploy higher-end Serial-Attached SCSI hard drives, whereas lower-cost Serial ATA drives can be appropriate in some smaller applications. Solid-state disks are also becoming more common, but they have a high price premium that might make them less attractive for storage-demanding video surveillance applications. To ensure redundancy, redundant array of independent disk arrays are often used. In some network video installations, network-attached storage (NAS) or storage area networks (SANs) are used.

In small installations of up to 50 cameras, a regular server would be appropriate. In even small systems of up to 16 cameras, edge storage can be used, such as a NAS solution or flash memory cards inside the cameras. In an edge storage solution, the cameras manage the recordings, which eliminates the need for external servers or storage.

Today, virtually any size storage system can be built. So to successfully install a network video system, it is critical to know what technologies are available, what level of performance and redundancy can be provided, and at what cost. For more information about servers and storage, see Chapter 12.

## 2.6 VIDEO MANAGEMENT SOFTWARE

---

For very small systems, a standard web browser that uses the built-in web interface offers adequate video management functions. This is at least true as long as only the video streams of one or a few cameras or video encoders are viewed at the same time and no recording is needed.

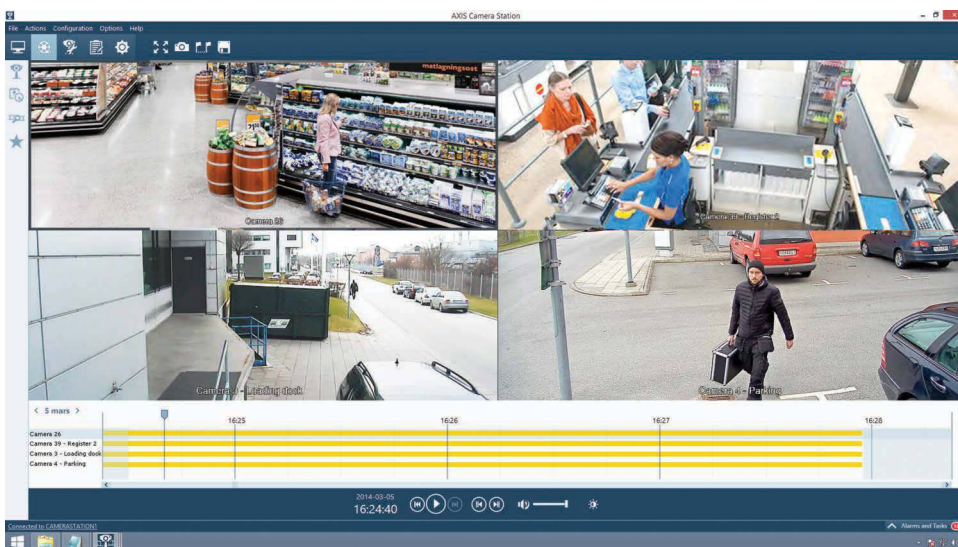
However, dedicated video management software is needed to effectively view several video streams at the same time.

Video management software is at the center of video management, monitoring, analysis, and recording (Figure 2.6). There are hundreds of different video management applications available from different companies around the world. They run on different operating systems (Windows, UNIX, Linux, and Mac OS®) and cover all vertical markets, multiple languages, and a range of scalability requirements. There are also numerous integration possibilities with other systems, such as building management, access control, and industrial control. Open platform solutions run on off-the-shelf hardware, which has components selected for optimal price and performance.

Manufacturers of network cameras and video encoders publish application programming interfaces (APIs) to ensure integration and compatibility with video management software. The more open the API, the better the collection of video management software and the tighter the integration is for a given network camera or video encoder. Ideally, the API should be published openly on the web and free of charge. There are also established standards, such as ONVIF, for interfacing to IP cameras and encoders.

Some of the common features provided by most video management software include the following:

- *Simultaneous viewing and recording of live video from multiple cameras:* Video management enables multiple users to view several different cameras at the same time and allows recordings to take place simultaneously.
- *Several recording modes:* Continuous, manual, scheduled, alarm-activated, and on motion and audio detection. Video motion detection defines activity by analyzing data and differences in a series of images. Video motion detection can be performed by the camera, which is preferable, or by the software.
- *Multiple search functions for recorded events:* Areas of interest can be defined and video from several cameras can be played and replayed at the same time.
- *Camera management:* Video management systems allow users to administer and manage cameras from a single interface. This is useful for tasks such as detecting cameras on the network, managing IP addresses, and setting resolution, compression, and security levels.
- *Remote access:* Control through a web browser, client software, or smartphone or tablet platform.



**Figure 2.6** Screenshot of a video management software interface.



- *Control of PTZ and panoramic cameras:* It can be controlled through a joystick or mouse by an operator, or it can be controlled through guard tours, which are automatic PTZ movement patterns set in the software or in the cameras.
- *Configuration of I/Os:* Enables video to be sent and recorded, external devices such as relays and lights to be automatically activated, and alarms to be sent in response to external sensors. This allows remote monitoring stations to become immediately aware of a change in a monitored environment.
- *Alarm management:* The software can activate an alarm, display pop-up windows, send emails, or send text messages (SMS) to cell phones.
- *Audio support:* Real-time support for either live or recorded audio.

Video management systems are easily scalable because cameras can be added one at a time, and some systems can scale to thousands of cameras. Open systems are suitable for scenarios where large numbers of cameras are used. They also make it easier to add functionalities to the system, such as increased or external storage, firewalls, virus protection, and intelligent video algorithms. Some video management systems use a web interface that makes it possible to access the video from any type of computer platform. When using the proper safeguards (e.g., password protection, security layering, and IP address filtering), web interfaces allow secure video online management from anywhere in the world.

Video management systems based on open platforms have another advantage in that they can be more easily integrated with other systems such as access control, building management, and industrial control. This allows users to manage video and other building controls through a single program and interface. Integrating a video surveillance system with an access control system, for example, allows for capturing video at all entrance and exit points, enabling photos in a badge system to be matched against images of the person actually using the access card.

To read more about video management software, see Chapter 13.

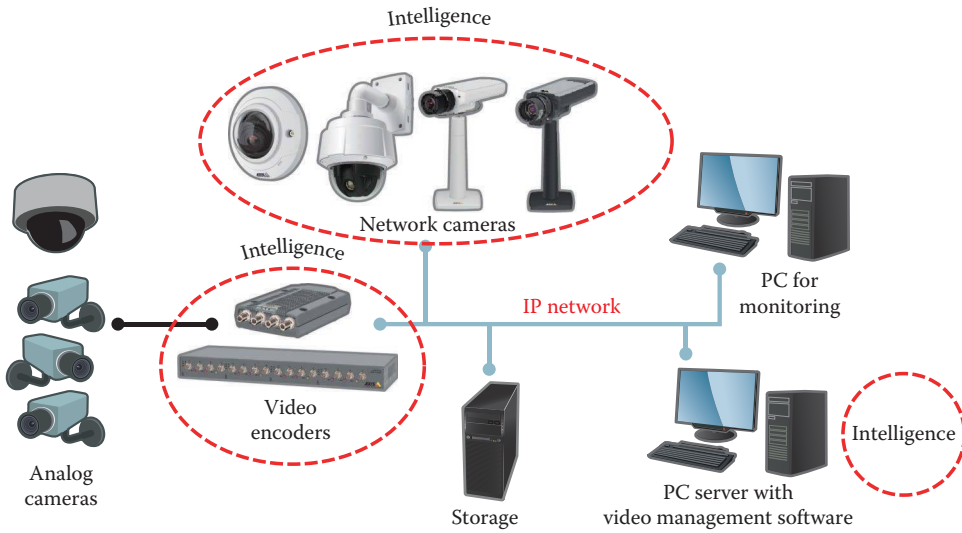
With the emergence of cloud-based computing, video management can now be offered as a service. This means that the cameras automatically record and make live or recorded video available in the cloud, which provides increased scalability for small systems and increased security by recording video off-site. For information about hosted video, see Chapter 14.

## 2.7 INTELLIGENT VIDEO

Intelligent video (Figure 2.7), also called video intelligence, video content analysis, or video analytics, is a market driven by a desire to increase the value of a video surveillance system and make it more proactive. Intelligent video makes response times faster, needs fewer operators per camera, and can extract information for uses beyond traditional surveillance, such as people counting. Research from Sandia National Laboratories in the United States has shown that an operator is likely to miss important information after only 20 minutes in front of a monitor. Intelligent video can help operators cover more cameras and respond more quickly.

The architecture of an intelligent video system can be either centralized or distributed. In a centralized system, all intelligence resides in the video management system. The main benefit of a centralized system is that the server is an open platform in most cases, so adding functionality is simple. Intelligent video algorithms consume a lot of computing power, which limits the number of cameras that can be managed by one server.

In a distributed system, the intelligence is distributed to the edge and resides in the network camera or video encoder. The main benefit of this is that the analysis is more accurate because uncompressed video is available in the camera. It also makes the system fully scalable and potentially enables reduced bandwidth usage because the camera or video encoder can determine, based on the video content, whether to send video to the server. Today, intelligent video is mostly used in



**Figure 2.7** A typical system using intelligent video. The intelligence usually resides at the edge, that is, built into the software or uploaded to the network cameras or the video encoders. It can also reside centrally, in the video management system.

high-risk transportation, government, and retail applications. The functionality includes camera tampering alarms, people counting, license plate recognition, motion detection, and object tracking.

Video intelligence is necessary in systems that require fast response times or where a very large number of cameras need to be proactively managed. For more information about intelligent video, see Chapters 15 and 16.



# Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>



## CHAPTER 3

### Network cameras

Network cameras, or IP cameras, offer a wide variety of features and capabilities to meet the requirements of almost any surveillance system. Today, network cameras offer improved image quality, higher resolution, and built-in storage and intelligence.

This chapter presents the components of a network camera and the different camera types. It also gives in-depth information about pan, tilt, and zoom (PTZ) network cameras, panoramic cameras, and day-and-night cameras, as well as megapixel, HDTV, and Ultra-HD network cameras. Additional information about camera technologies is provided in Chapter 4.

#### 3.1 NETWORK CAMERA COMPONENTS

---

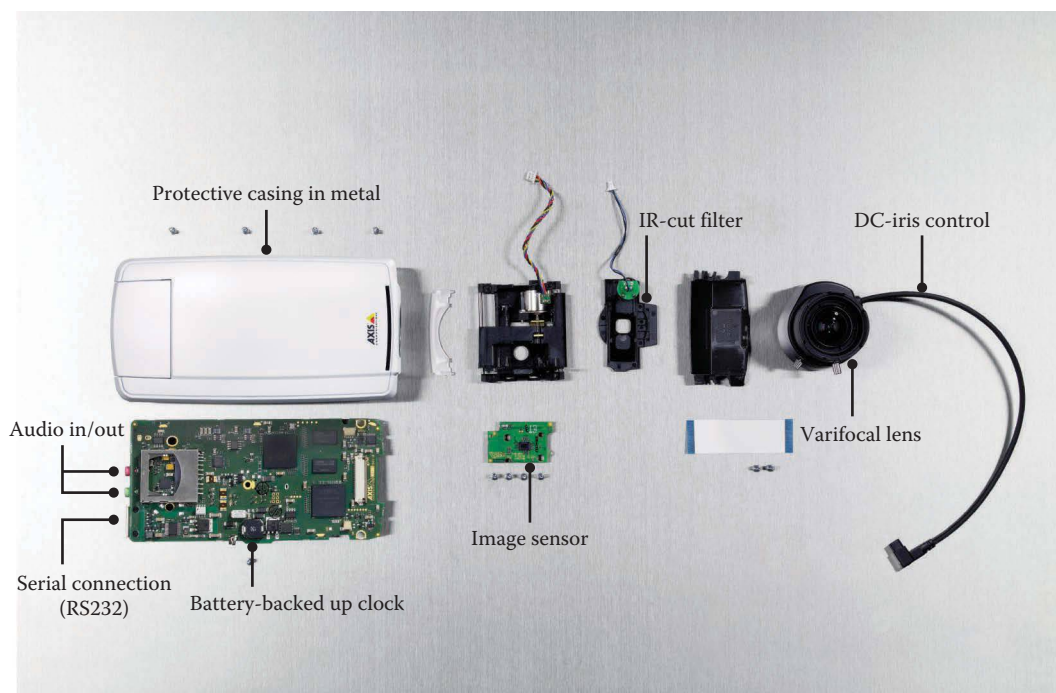
As the name implies, a network camera is a camera with a network connection. The network camera can be described as a combined camera and computer. Like a computer, the network camera has its own IP address, is connected directly to a network, can be placed wherever there is a network connection, and can send live and recorded video and audio over the network.

The main components of a network camera (see Figure 3.1) include the following:

- Lens for focusing the image on the image sensor
- Image sensor, either a charge-coupled device or a complementary metal-oxide semiconductor
- Processor, one or several, for image processing, compression, video analysis, and networking functionality
- Flash memory for storing the camera's firmware code
- Random access memory and/or SD card for local recording of video clips and events

In the analog world, cameras exist primarily in two formats: phase alternating line (PAL) (the European CCTV standard) and national television system committee (NTSC) (the American CCTV standard). In the network video world, both PAL and NTSC are irrelevant because network cameras use globally established resolution standards such as VGA or HDTV 1080p. Network cameras connect to Ethernet cables, which transport video that is usually compressed in compliance with worldwide standards such as Motion JPEG and MPEG-4 (H.264). This makes planning and servicing worldwide video surveillance deployments much easier. For more information about video compression formats, see Chapter 6.

Most network cameras also offer input/output ports that enable connections to external devices such as sensors and relays. Built-in support for Power over Ethernet (PoE) eliminates the need for separate power cabling to the camera. Another time and money saver in the installation process is the ability to remotely set the field of view and focus the camera.



**Figure 3.1** Components of a network camera.

A network camera can have many advanced features, such as built-in intelligence, audio monitoring, and alarm handling. The networking functionality is also essential and should include all the latest security and IP protocols. Different cameras offer different performance levels. Some cameras provide less than full frame rate of video, which is considered to be 30 fps, whereas others can provide 60 fps or several simultaneous video streams at full frame rate.

## 3.2 TYPES OF NETWORK CAMERAS

Network cameras can be classified in terms of whether they are designed only for indoor use or for both indoor and outdoor use. Outdoor network cameras often have an auto-iris lens to regulate the amount of light to which the image sensor is exposed. An outdoor camera also requires an external protective housing unless the camera design already incorporates a protective enclosure. Housings are also available for indoor cameras that need protection in harsh environments, for example, dusty or humid conditions, or in areas where there tend to be a lot of vandalism or tampering attempts. Most modern camera types have a protective enclosure that is appropriate for the intended use, such as an indoor, outdoor, vandal-resistant, or dust-proof housing. For more information about enclosures, see Chapter 17.

More commonly, cameras are sorted into groups based on their view functionality: fixed cameras, fixed dome cameras, PTZ cameras, and panoramic cameras. There are also niche cameras for special applications like covert and thermal cameras. All these types are described in the following sections.

### 3.2.1 Fixed cameras

A fixed camera has a fixed viewing direction. When mounting the camera, the installer sets the view and sometimes the focus. With this traditional camera type, the camera and the direction in which it is pointing are clearly visible. Therefore, it is the best camera choice in situations where it is advantageous that the camera is very noticeable. A fixed camera may come with a fixed, varifocal,



**Figure 3.2** A fixed camera.

or motorized zoom lens. Most fixed cameras have exchangeable lenses, which are convenient and cost effective because the lens can be switched if the requirements change. Fixed cameras can be installed in protective enclosures and can also be mounted on a pan-tilt motor for greater viewing flexibility. Figure 3.2 shows an example of a fixed camera.

### 3.2.2 Fixed dome cameras

Fixed dome cameras, also called minidomes, consist of a fixed camera preinstalled in a small dome-shaped housing. The camera can be easily adjusted to point in any direction. This camera's main benefit lies in its discreet, unobtrusive design. In addition, people in the camera's range find it difficult to see the direction in which the camera is pointing. The fixed dome camera is also more resistant to tampering than a fixed camera. One of the limitations of a fixed dome camera is that it rarely comes with an exchangeable lens. Even if the lens is exchangeable, the choice of lenses is limited by the space inside the dome housing. However, a varifocal lens is often available, which enables adjustment of the camera's field of view either manually or, if the lens is motorized, remotely. Fixed dome cameras are always designed with different types and levels of protection, such as vandal and dust resistance, and IP66 and NEMA 4× ratings for outdoor installations. No external housing is required. These cameras are usually mounted on a wall, ceiling, or pole. Figure 3.3 shows an example of a fixed network camera.



**Figure 3.3** A fixed network camera.

### 3.2.3 PTZ cameras

PTZ cameras are designed with electrical motors that allow the camera to move in three dimensions. Through panning, tilting, and zooming, the camera has the ability to adjust the scene that is monitored. For example, instead of following a person in a retail store in person, operators can follow the person through a PTZ camera that they control manually with a joystick.

Through their pan, tilt, and zoom functionalities, PTZ cameras can cover a wide area. How much PTZ cameras can move differs between types and models. The most advanced models offer a full 360° pan angle and usually a tilt angle of 180°, but sometimes even more so they can “look above the horizon.” Figure 3.4 shows an example of a PTZ camera.

Available at a lower price are types of PTZ cameras that are not designed for full 360° panning (Figure 3.5). Other types have very few moving components because their designers avoid using



**Figure 3.4** A pan, tilt, and zoom camera.



**Figure 3.5** A pan, tilt, and zoom camera that lacks full panning capability. It is still useful in many surveillance scenarios and generally comes at a lower cost.



**Figure 3.6** A wall-mounted pan, tilt, and zoom camera.

belts that need to be serviced. These cameras typically have limited panning and are often designed to be mounted on a wall (Figure 3.6).

PTZ cameras are ideal in discreet installations thanks to their design. There are mounting options (as seen in Figure 3.7) that allow the cameras to blend in. These mounts are known as drop-ceiling mounts, recessed mounts, or flush mounts. The domes (domes can be clear or smoked) make it difficult for people to identify the camera's viewing angle. A high-end PTZ camera also has the mechanical robustness that is needed to run in guard tour mode. In guard tour mode, the camera moves continuously between presets, and this allows one PTZ camera to cover an area where several fixed cameras would otherwise be needed. The main drawback is that only one location can be monitored at any given time, leaving the other positions unmonitored.

A PTZ camera's optical zoom typically ranges between 3× and 40×. It is often used in situations where an operator is present and can control it. This type of camera is usually mounted on a ceiling if used indoors or on a pole or corner of a building in outdoor installations. A PTZ camera gets all its PTZ commands over an IP network, so no RS485 wires need to be installed unlike an analog PTZ camera. Also, most PTZ cameras can be powered using PoE instead of a dedicated power source. Both these factors help lower the total installation cost. (Read more about PTZ cameras in Section 3.3.)



**Figure 3.7** A pan, tilt, and zoom camera in a drop-ceiling mount. When installed, only the trim ring and the dome-covered lens assembly are visible below the ceiling.

### 3.2.4 Panoramic cameras

A panoramic camera has the ability to deliver 360° or 180° surveillance, providing full overview images as well as close-up images without moving camera parts. This is achieved by taking advantage of the high resolution provided by a megapixel sensor and a wide-angle lens. When zooming in for close-ups, the camera instantly switches field of view. It can also provide multiple streams of video from different field of views. Figure 3.8 shows an example of a panoramic camera. (Read more about panoramic cameras in Section 3.4.)

### 3.2.5 Covert cameras

Covert cameras are designed to blend into the environment and be virtually impossible to discover. Covert analog cameras, using the limited PAL/NTSC resolutions, have existed for a long time and have often been used together with network video encoders. (For more information about encoders, see Chapter 8.) To bestow all the benefits of network video in a covert application while making the camera as discreet as possible, the sensor unit has a high-resolution (HDTV) sensor that is often decoupled from the central unit that contains the processor. The cable between the sensor unit and central unit is typically less than 10 m (33 ft). In many cases, you can combine different central and sensor units in several different ways. Therefore, covert cameras are sometimes referred to as modular cameras. Figure 3.9 shows an example of a covert camera.



**Figure 3.8** A panoramic camera.



**Figure 3.9** A covert camera. A cable connects the sensor unit to the central unit (also known as main unit).

The perfect scenario and placement of these cameras is at eye level in a retail store entrance or flush mounted in an ATM machine environment to provide discreet or covert surveillance. They can offer general overview surveillance or close-up shots for identification. Because they are so small and sometimes nearly invisible, they are subject to very few tampering attempts.

### 3.2.6 Thermal cameras

Thermal cameras create images based on heat that radiates from all objects. Most thermal cameras come in the shape of a fixed camera. Figure 3.10 shows an example of a thermal camera. Images are generally produced in black and white but can be artificially colored to make it easier to distinguish different shades. A thermal camera can see what a standard network camera that works in the visible and near-infrared (NIR) range cannot (see Figure 3.11). The contrast in thermal images is higher when there are great temperature differences in a scene.

Thermal cameras are ideal for detecting people, objects, and incidents in shadows, in complete darkness, or in other challenging conditions such as smoke and dust. Because thermal images do not enable reliable identification, thermal cameras are used primarily to detect suspicious activities. They complement and support conventional network cameras in a surveillance installation by



**Figure 3.10** An outdoor thermal camera.



(a)



(b)

**Figure 3.11** The same scene as seen by a standard network camera (a) and a thermal network camera (b).



providing perimeter or area protection, discreet surveillance, and security in dangerous or off-limit areas such as tunnels, railway tracks, and bridges.

For detailed information about thermal cameras, see Chapter 5.

### 3.3 PTZ CAMERAS

PTZ cameras provide several unique benefits, and network PTZs in particular have increased the capabilities of the cameras by increasing resolution, providing better image quality, as well as local storage and built-in intelligence. Because the same network cable can be used for transporting video, sending PTZ commands, and supplying power to the camera, it is more cost efficient to install PTZ cameras than it was in the past.

One advantage of some of the latest PTZ cameras is the use of high-resolution (HDTV) sensors. Because these sensors use progressive scan, there is less motion blur in the images they produce. For moving cameras, this is particularly beneficial because their movements add blurriness to their images if the traditional interlace scanning technology is used. For more information about interlaced and progressive scanning, see Chapter 4.

This section covers some of the specific features found only in PTZ cameras, such as image stabilization, guard tours, privacy masking, auto-flip, and E-flip.

#### 3.3.1 Image stabilization

In outdoor installations, zoom factors above 20× could be impractical due to vibrations and motion caused by traffic or wind. Vibrations in a video can be reduced if the camera is equipped with an image stabilizer. An image stabilizer is especially useful in windy environments such as harbors or areas prone to heavy vibrations such as highways or bridges. Even in office buildings, there is normally some level of vibration.

There are several types of image stabilization. Mechanical image stabilization or optical image stabilization changes the path to the sensor by moving the sensor or the lens. Image stabilization can also be made within the camera's processor. This is called electronic image stabilization (EIS) and is achieved through a combination of motion estimation and image shift. EIS generally requires a slightly oversized sensor and means that the field of view of the output image is smaller than the actual sensor image.

Blurring that is caused by very rapid movements that occur within the exposure time of the camera can be reduced by decreasing the exposure time. It is also possible to use special lenses that compensate motion during the exposure time.

EIS also reduces the file size of the compressed image and lowers the bitrate of the video stream. In addition to the obvious benefit of getting video that is more useful, using EIS saves valuable storage space (see Figure 3.12). Historically, EIS has primarily been available in PTZ cameras but can also be found in some high-end fixed cameras.

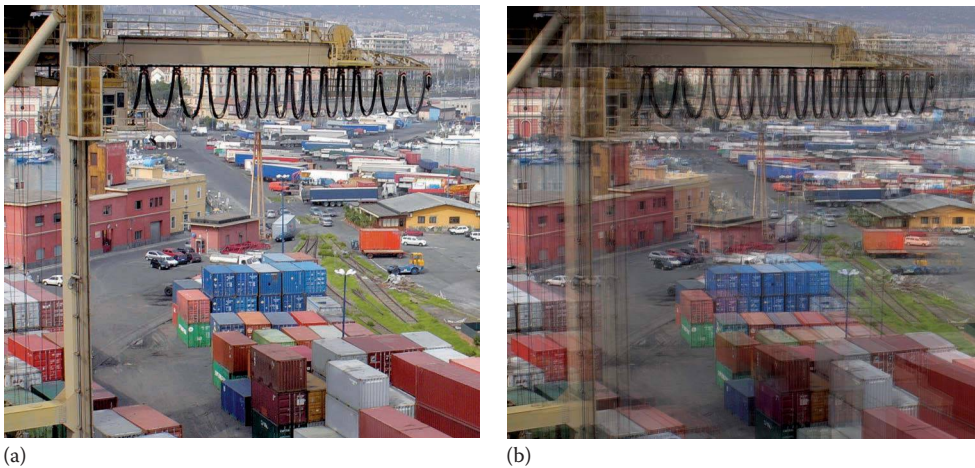
#### 3.3.2 Presets and guard tours

Many dome and PTZ cameras have a number of preset positions, often 100 or more (Figure 3.13). Once the preset positions have been set in the camera, it is quick and easy for the operator to go from one position to the next. Many PTZ cameras also have built-in guard tours (Figure 3.14). A guard tour enables the camera to move automatically from one preset position to the next in a pre-determined or random order. The viewing time between one position and the next is configurable. Different guard tours are active during different times of the day.

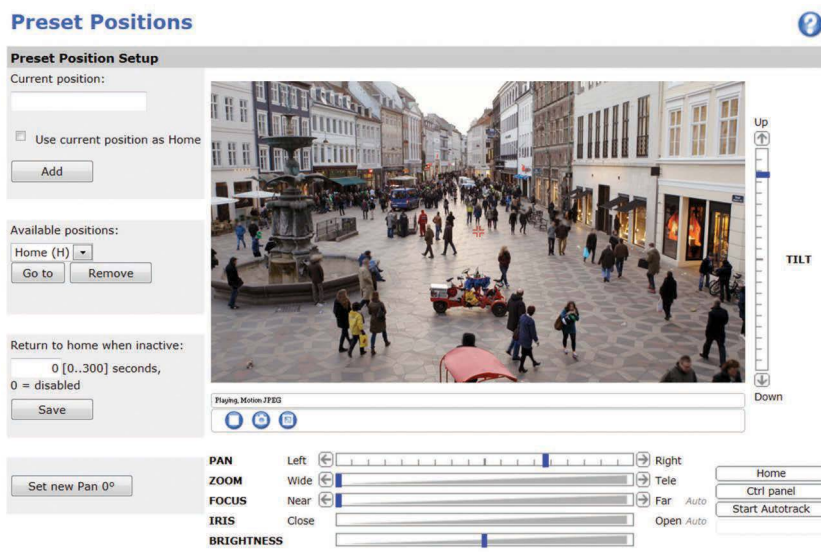
#### 3.3.3 Privacy masking

A PTZ camera can provide surveillance of very large areas, sometimes even in areas where the operator should not monitor. For example, a PTZ camera installed outside a football stadium





**Figure 3.12** A snapshot of video with electronic image stabilization (EIS) (a) and without EIS (b).



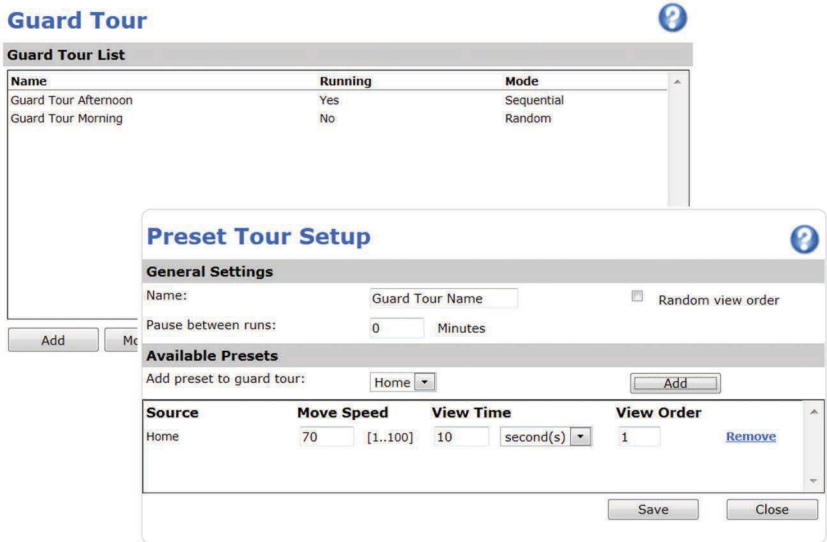
**Figure 3.13** A typical interface for preset positions. Some cameras offer as many as 256 preset positions.

in a city center may likely be able to view and zoom in on a nearby apartment complex. In such instances, privacy masking becomes important because it allows some areas of a scene to be blocked or masked from viewing and recording (Figure 3.15).

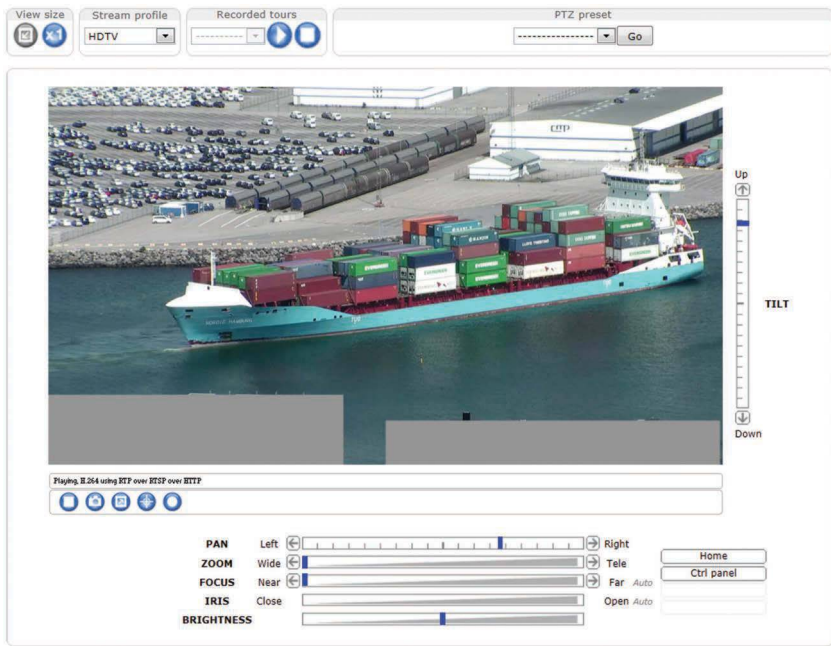
Cameras with digital PTZ also have privacy mask functionality. But in cameras with mechanical PTZ capabilities, the privacy mask does not only follow the size and position of the object but also shifts angle, grows, and shrinks with the object.

### 3.3.4 E-flip

Imagine a scenario similar to the following example. A PTZ camera is mounted in a ceiling in a retail store and is tracking a person who is suspected of shoplifting. Then, the suspect passes just underneath the camera. A camera with E-flip electronically and automatically rotates the images  $180^\circ$  so that the suspect and the rest of the image keep having the right side up. Without E-flip, the whole image would be upside down when the suspect passes underneath the camera. The E-flip function is performed automatically and will not be noticed by an operator.



**Figure 3.14** An example of interface and dialog for guard tour setup. In many cases up to 20 guard tours can be programmed.



**Figure 3.15** An example of privacy masking. Note the gray boxes at the lower edge of the image.

### 3.3.5 Auto-flip

Slip ring is the mechanical component that gives the camera the ability to do full 360° panning. Some PTZ cameras do not have slip rings because of cost or size limitations. Those PTZ cameras have a mechanical stop that prevents 360° panning. Some of these cameras can bypass the mechanical stop by automatically freezing the image for a second while panning 180° and simultaneously tilting the camera block to the previous position. Then, the camera continues the smooth panning that was started by the operator. These movements are illustrated in Figure 3.16.



**Figure 3.16** A camera with auto-flip can deliver 360° panning even if it has a mechanical stop. The camera pans 180°, tilts to its starting position, and then continues the panning operation.

### 3.3.6 PTZ performance

The mechanical performance levels of PTZ can differ a lot. Datasheets often specify the maximum performance in degrees per second. If the stated performance of a PTZ camera is more than 360° per second, it means that the camera can complete a full circle in less than 1 second. Normally, this is considered adequate for a high-performance PTZ camera. Some PTZ cameras can move even faster—up to around 450° per second. Although speed is important, the camera's movements also need to be controllable at very low speeds. Therefore, some PTZ cameras specify the slowest speed with which they can be controlled. For example, a speed of 0.05° per second means it will take 2 hours to complete just one 360° rotation.

### 3.3.7 Joystick control

Joysticks make it very easy to control a PTZ camera. USB joysticks, which are connected to the PC used for video monitoring, are commonly used in network video applications. As mentioned earlier, most professional joysticks also come with buttons that can be used for presets. Figure 3.17 shows an example of a PTZ joystick.



**Figure 3.17** A pan, tilt, and zoom joystick.

### 3.4 PANORAMIC NETWORK CAMERAS

A panoramic camera can often be compared to a nonmechanical PTZ camera. It uses a high-resolution sensor and a wide-angle lens, which gives it a viewing angle of  $100^{\circ}$ – $180^{\circ}$  (or even wider in some cases). Panoramic cameras are sometimes referred to as  $360^{\circ}$  cameras because they can monitor in any direction when they are mounted in a ceiling. From the outside, panoramic cameras (see Figure 3.8) often look like fixed dome cameras. The wide-angle lens and megapixel sensor offer multiple viewing modes such as  $360^{\circ}$  overview,  $180^{\circ}$  panorama, and quad views (simulating four different cameras) and view areas with support for digital PTZ functionality.

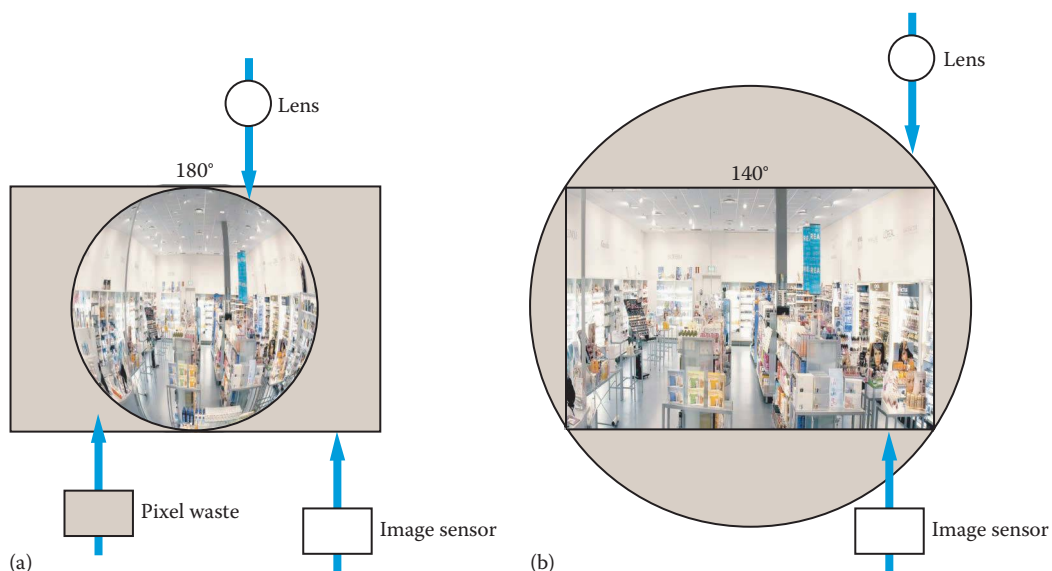
A panoramic camera allows the operator to zoom in on any part of a scene without mechanical movement. The key advantage is that there is no wear and tear because the camera has no moving parts. The PTZ movements are all digital. Zooming in on a new area of a scene is immediate. In a mechanical PTZ camera, zooming can take up to 1 second. Because a panoramic camera's viewing angle is not visible, it is ideal for discreet installations. The disadvantage is that the image detail is not anywhere close to that of a real PTZ camera when zooming in more than  $4\times$  or  $5\times$ . The higher the resolution of the sensor, the better the zooming ability, but it comes at the price of mediocre light sensitivity.

One type of panoramic camera uses multiple sensors to provide even higher resolution. An algorithm stitches the images together. Because the operator sees just one image, they can treat it as they would a single-sensor camera. This kind of camera delivers better image detail but commands a higher price.

#### 3.4.1 Selecting the right viewing angle

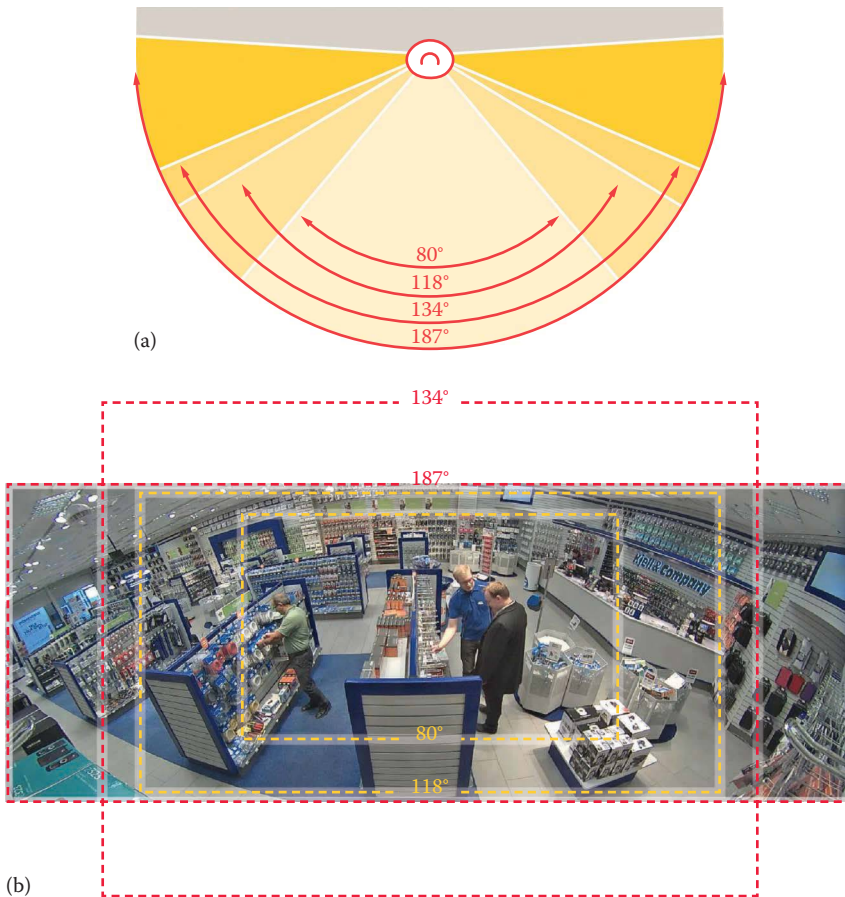
Depending on the type of lens used, a camera can have a viewing angle of up to  $180^{\circ}$ . A wide viewing angle, however, comes with some challenges. One challenge is that the image is greatly distorted by the fish-eye effect from the lens. Therefore, the image needs to be dewarped to make it viewable. Dewarping an image requires a lot of processing power in the camera or viewing station, which reduces the equipment's performance. The other problem is that only a very limited part of the image sensor can be used, which reduces resolution and thereby image quality (see Figure 3.18).

The images in Figure 3.19 illustrate the difference between different fields of view.



**Figure 3.18** With a  $180^{\circ}$  field of view, only part of the sensor is used to make the image (a). With a  $140^{\circ}$  field of view, all of the sensor's pixels are used to make the image (b).





**Figure 3.19** Different viewing angles have different fields of view: diagram of the different viewing angles (a). The resulting field of view of different viewing angles, where the 187° view is a panoramic view (b).

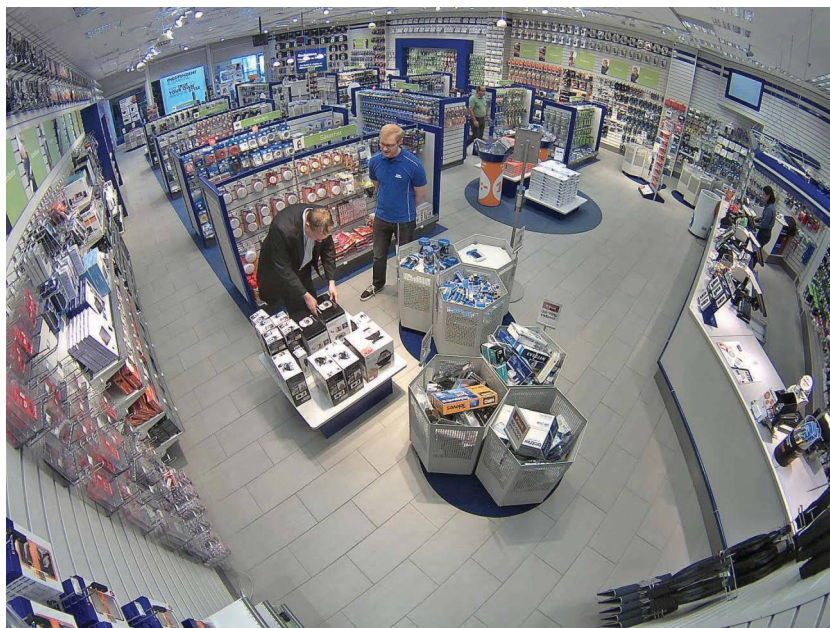
### 3.4.2 Cameras with wide viewing angles

Cameras with extra wide viewing angles, typically 120°–150°, are technically in between what is traditionally referred to as panoramic cameras and regular cameras. They offer surveillance for large areas (see an example in Figure 3.20). In this case, the horizontal and vertical angles of view are combined with a highly effective megapixel sensor to provide a sharp image with good detail, especially in the corners. When this kind of camera is housed in vandal-resistant casing, it is ideal for effective surveillance in public transport terminals, stores and malls, hotels, banks, and school areas.

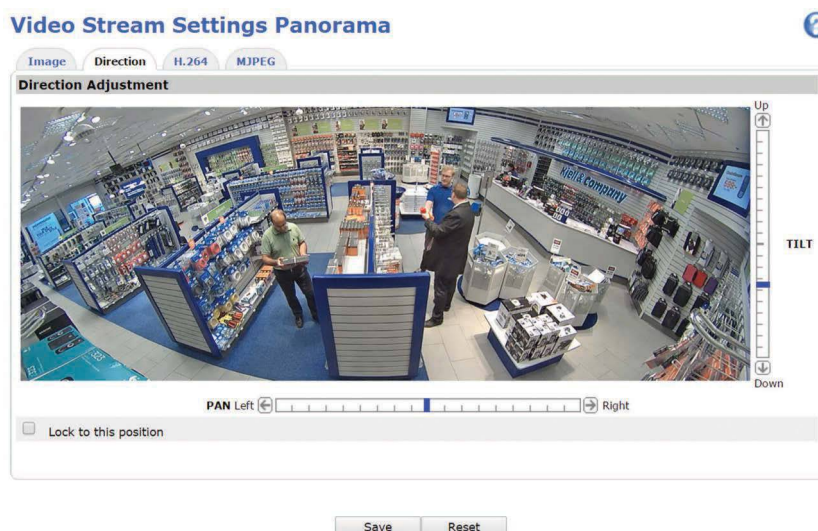
The camera in this example supports digital PTZ, which can be used to simulate a varifocal lens to adjust the field of view remotely after the camera's physical installation. The camera's digital PTZ and multiview streaming abilities allow different areas of a scene to be cropped from the full view. This transforms the camera into several virtual cameras and helps minimize bitrate and storage needs. The camera can send multiple, simultaneous streams at different compression rates. For example, one stream can be used for viewing and another for recording.

### 3.4.3 180° panoramic cameras

A 180° panoramic camera provides panoramic views in a high-resolution image. The camera can be mounted on a wall or ceiling, and the example in Figure 3.21 shows an image from a wall-mounted, fixed minidome that also offers digital PTZ and multiview streaming with dewarped views. It is ideal for public areas where event monitoring such as motion detection is required.



**Figure 3.20** A corner view from a panoramic camera with a 134° viewing angle.

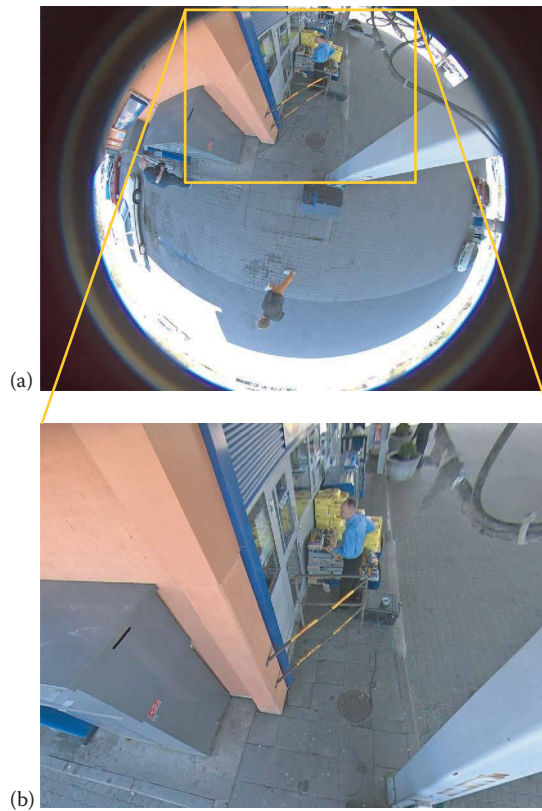


**Figure 3.21** An example of a panorama view in a settings window.

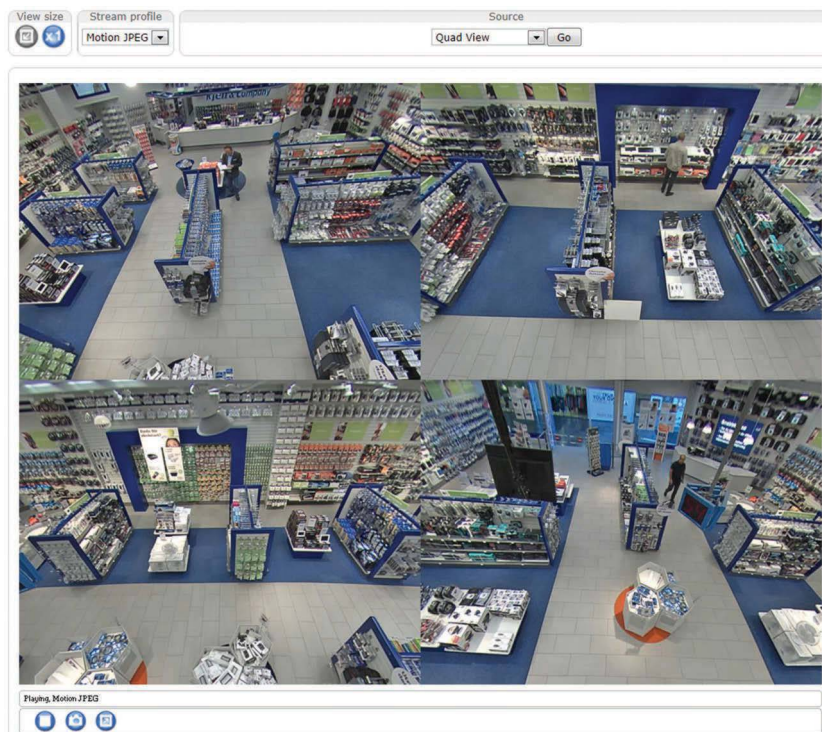
### 3.4.4 360° panoramic cameras

When mounted in a ceiling, a 360° panoramic camera provides a variety of views. Figure 3.22 shows images for the same scene from a ceiling-mounted panoramic camera. Different aspects of the scene can be viewed from different perspectives.

This kind of camera can also offer quad view, which is helpful, for example, when the camera is positioned at the intersection of two aisles. Users can digitally pan, tilt, and zoom in on areas of interest in four individually cropped out and dewarped view areas (see Figure 3.23).



**Figure 3.22** An overview image (a) and the results from the same scene after dewarping (b).



**Figure 3.23** An example of a quad view from a ceiling-mounted panoramic camera.



### 3.4.5 Multisensor panoramic cameras

Multisensor systems are gaining ground in the surveillance industry. These camera systems use multiple sensors that are paired with individual lenses. This combination creates an omnidirectional system that provides panoramic images without distortion. Figure 3.24 shows an example of a multisensor panoramic camera.

A multisensor camera typically includes three or four individual cameras combined into one unit and is therefore more expensive than wide-angle panoramic cameras with just one camera. Also, images from multisensor cameras usually overlap and are often stitched together to make a complete image. Figure 3.25 shows an example of a stitched image from a multisensor camera. Although this is a complex technology, demand and development are rapidly bringing multisensor cameras into the market.

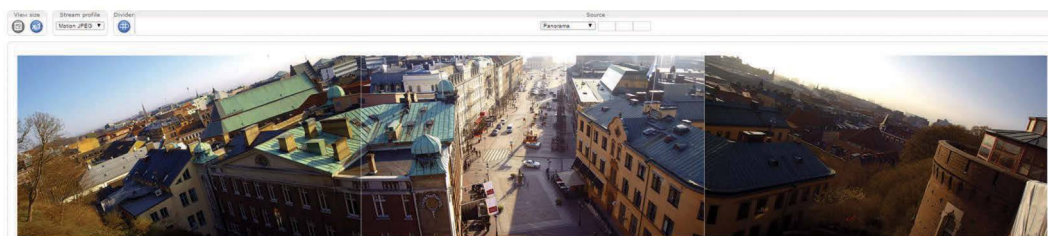
### 3.4.6 Comparing and combining panoramic and PTZ cameras

A panoramic network camera offers a wider field of view than a mechanical PTZ camera, and there is no wear and tear caused by moving parts. The PTZ camera delivers much better image detail, especially for objects that are far away. The typical application for a panoramic camera is for situational awareness—to detect *whether* a specific activity is occurring. To identify *what* happened normally requires many fixed or fixed dome cameras along with one or more PTZ cameras. The images in Figure 3.26 illustrate the difference in field of view between panoramic and PTZ cameras.

Sometimes, the choice between overview and detail becomes impossible. Perhaps, a security manager wants the best of both worlds without having to spend money or time on setting up several

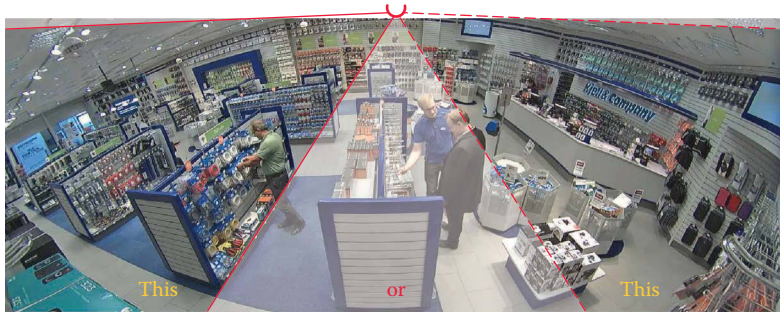


**Figure 3.24** A multisensor camera.



**Figure 3.25** An example of a stitched image from a multisensor camera.





(a)



(b)

**Figure 3.26** Pan, tilt, and zoom cameras can cover bigger areas and can see details from far away, but they can only see one section a time (a). Cameras with wide and panoramic viewing angles give an overview of the whole area (b).

different cameras to monitor a single large area. The solution is a combination camera that includes a multisensor camera and a PTZ camera. All its integrated cameras can be powered and controlled through the same network cable. This type of camera is sometimes referred to as an omnicamera. It has a permanent panoramic overview as well as the ability to focus in on an area and follow a subject once an activity has been detected. It is an ideal camera for many situations and areas such as city squares, stadiums, parking lots, airports, and logistic centers. Figure 3.27 shows an example of this type of combination camera.



**Figure 3.27** A multisensor and pan, tilt, and zoom (PTZ) camera combination. This particular multisensor camera has four sensors and lenses for monitoring a larger area. It also controls the integrated PTZ camera. This solution allows the operator to keep the big picture and zoom in on details at the same time.



**Figure 3.28** An onboard camera mounted in the ceiling of a railroad passenger car.

### 3.5 ONBOARD CAMERAS

Onboard cameras are specially designed for discreet and efficient surveillance in trains, subway cars, buses, and emergency vehicles. The circumstances of onboard surveillance demand a particularly rugged design. The cameras need to be dust and water protected and withstand tough conditions such as vibrations, shocks, bumps, and temperature fluctuations. The modest appearance, which is usually based on a fixed minidome, is often combined with an active tampering alarm that helps detecting and preventing tampering attempts such as redirection and defocusing. Figure 3.28 shows an example of an onboard camera in its intended environment.

### 3.6 DAY-AND-NIGHT NETWORK CAMERAS

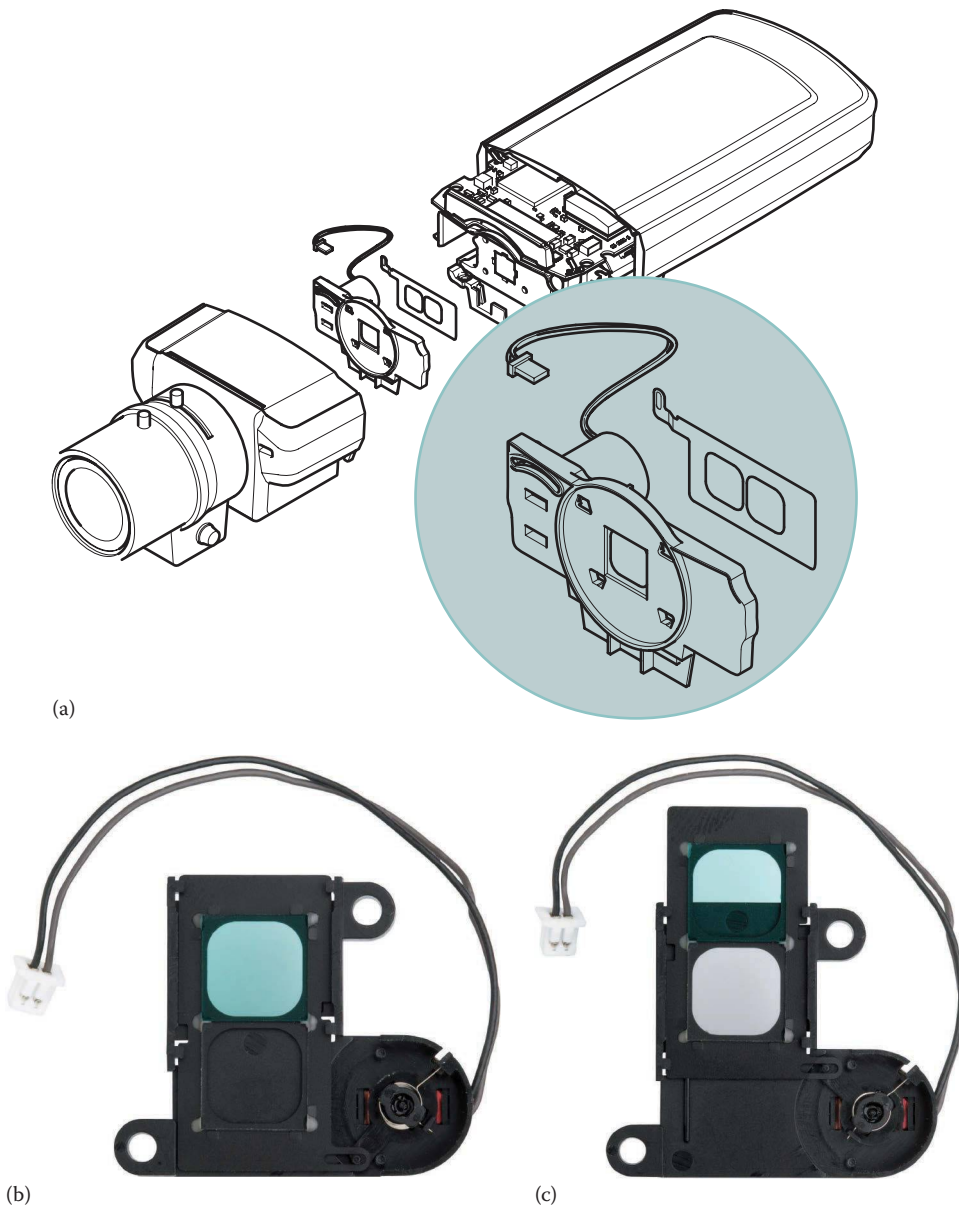
All types of cameras (such as fixed cameras, fixed dome cameras, and PTZ cameras) can offer day-and-night functionality. A day-and-night camera is designed for use in outdoor installations or in indoor environments with poor or no lighting. The day-and-night functionality can be achieved in two different ways. The simple way is to turn it into a black-and-white camera by decreasing the chrominance (color) sensitivity of the camera, which somewhat improves the camera's light sensitivity. However, this is not true day-and-night functionality although some vendors market it as such.

True day-and-night functionality is only achieved if a mechanically removable IR cut filter, also called an IR-blocking filter, is used in the camera. When the filter is removed, the image sensor can pick up light in the NIR wavelength range. This increases the camera's light sensitivity so it can reach down to 0.001 lux or lower. In most true day-and-night cameras, the IR cut filter is removed automatically when the light that hits the sensor is below a certain level, for example, below 1 lux (Figure 3.29).

A day-and-night network camera delivers color images during the day. As the light dims, the camera switches to night mode and delivers video in black and white to reduce noise (graininess) and to provide clear, high-quality images (Figure 3.30).

#### 3.6.1 IR illuminators

IR illuminators should be used if the goal is to conduct 24/7 surveillance in areas with low light. Many different types of IR illuminators are available. Some illuminators are external and stand alone, some can be mounted on cameras or their mounts, and some cameras come with built-in IR illuminators (see the images in Figure 3.31). Some IR illuminators produce light at a low



**Figure 3.29** An IR cut filter in a day-and-night network camera (a). The position of the IR cut filter during daytime (b). The position of the IR cut filter during nighttime (c).

wavelength, about 850 nm, which will display a faint red hue that is visible to the human eye if the room is completely dark. Totally covert IR illuminators with higher wavelengths of about 950 nm also are available. The downside of higher wavelengths is that they reach a much shorter distance. Illuminators also come in different illumination angles, for example, 20°, 30°, or 60°. A wide-angle illuminator normally equals a shorter reach. Most IR illuminators today use light-emitting diodes (LEDs), which is a cost-efficient solution that enables the illuminators to last between 5 and 10 years. New LED technology has made it possible to produce efficient solutions where only a few LEDs are mounted around the camera. The lifespan of these LEDs is very long, often longer than the camera's. Modern built-in LEDs also adjust the angle of the IR light as the viewing angle of the camera changes along with the camera changing the viewing angle. Figure 3.32 shows snapshots of a scene without and with IR illuminators.



(a)



(b)

**Figure 3.30** Image of a scene taken by a day-and-night camera in day mode (a) and night mode (b).

### 3.6.2 Day-and-night applications

Day-and-night or IR-sensitive cameras are useful in certain environments or situations that restrict the use of artificial light. They include low-light video surveillance applications, where light conditions are less than optimal, covert surveillance situations, and discreet installations, for example, in residential areas where bright lights could disturb residents at night. An IR illuminator that provides NIR light can also be used together with a day-and-night or IR-sensitive camera to further enhance the camera's ability to produce high-quality video in low-light or nighttime conditions.

With recent improvements in sensor and processing technology, network cameras not only are becoming more light sensitive but also are able to produce images with color at night. In some low-light scenarios, these network cameras are an even better option than day-and-night cameras.

## 3.7 MEGAPIXEL NETWORK CAMERAS

In the early years of megapixel network cameras, one of the most highlighted benefits was their ability to provide video with much higher resolution than analog cameras could ever deliver. Megapixel network cameras incorporate an image sensor that delivers images with a million or more pixels. However, with the fast adoption of high-resolution network cameras, nobody is





(a)

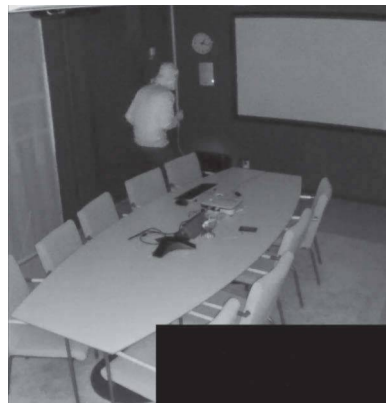


(b)

**Figure 3.31** IR illuminators (a) and a camera with built-in IR illuminators (b).



(a)



(b)

**Figure 3.32** Image of a scene taken by a camera with built-in IR illuminators. The IR illumination is off (a) and the IR illumination is on (b).

comparing megapixels against analog and VGA resolutions anymore. The market has matured, and there is very little demand for cameras below 720p (0.9 megapixels). In fact, many surveillance managers look for higher resolutions than that, and it is not uncommon that they need 5 megapixels or more. These cameras are sometimes referred to as multimegapixel or high-resolution cameras. Users also look for cameras that follow the HDTV standards 720p and 1080p found in flat-screen TVs. There has been a tradition that the demand for resolutions follows the consumer television market. And now that Ultra-HD TVs have been embraced by early-adopter consumers, the first 4K Ultra-HD network cameras have been introduced, providing a resolution of  $3840 \times 2160$  or 8 megapixels. For more information about megapixel, HDTV, and Ultra-HD resolutions, see Chapter 4.

### 3.7.1 Benefits of megapixel

A megapixel fixed camera can be used in two ways: it can enable viewers to see greater details (having higher resolution) in an image and it can be used to cover a larger part of a scene if the image scale (in pixels per area) is kept the same as a nonmegapixel camera. The advantages of the megapixel camera include easier identification of both people and objects and improved coverage of the most important areas of a scene, which ultimately leads to bandwidth and storage savings. In addition, one megapixel camera can serve in the place of multiple analog cameras and thereby reduce installation, operation, and maintenance costs (see Figures 3.33 and 3.34).

### 3.7.2 Megapixel applications

All camera types discussed in Section 3.2, with the exception of thermal cameras, are available in megapixel resolution. The great benefits of megapixels and the recent improvements in light sensitivity have led to rapid adoption of megapixel cameras. Today, more than 75% of network cameras are megapixel and some camera types are available only in high-resolution versions. Typically, fixed cameras have the highest resolutions—some have over 10 or 20 megapixels. Megapixel PTZ cameras are typically only available in 720p and 1080p.

Megapixel is useful in most applications, including the following:

- *Retail:* Effective video surveillance can help drastically reduce theft and shrinkage. Megapixel network cameras can either provide an overview of a large part of a store (without any blind spots) or offer highly detailed images of the area around the checkout counter.
- *City surveillance:* Megapixel network cameras provide high-resolution video streams from locations where it is necessary to conclusively identify people and objects or get a larger overview while viewing live or recorded video.



**Figure 3.33** The images from the analog camera (a) are not as sharp as those from a megapixel camera (b).



**Figure 3.34** The gray lines show the view from an analog camera, while the total image shows the megapixel camera view. Given that the number of pixels per area is the same, one 1080p (2 megapixels) camera covers a much larger scene than four 4CIF cameras.

- *Government buildings:* Megapixel network cameras provide the exceptional image detail necessary to facilitate the identification of people and to record evidence of any suspicious behavior.
- *School buildings:* The use of megapixel resolution cameras in hallways makes it easier to identify students.

The most common resolutions are HDTV 720p (0.9 megapixels) and HDTV 1080p (2 megapixels) along with 5 megapixels. The next logical step is for more cameras to be introduced with 4K Ultra-HD resolution (8.3 megapixels). (For more about resolutions, see Chapter 4.)

### 3.7.3 Drawbacks of megapixel

Although megapixel resolution enables either greater detail or more of a location to be seen, there are drawbacks. Typically, high-resolution megapixel sensors are the same size or only slightly larger than lower-resolution sensors. This means that while there are more pixels on a high-megapixel sensor, the size of each pixel is smaller than the size of each pixel on a lower-resolution sensor. The general rule is that the smaller the pixel size, the lower the light-gathering ability. So many high-resolution network cameras are less light sensitive than low-resolution network cameras. But through technical advancement, many modern 720p, 1080p, and 5-megapixel cameras have very good light sensitivity.

Not all high-megapixel cameras (5 megapixels and above) have the ability to provide 30 fps at full resolution. Another factor to consider when using megapixel network cameras is that higher-resolution video streams increase the demands on network bandwidth and storage space for recorded video. This can be somewhat mitigated using the H.264 video compression standard. (For more about compression, see Chapter 6.)

Another important consideration with megapixel cameras is the selection of lenses. The typical analog CCTV lens does not match the high resolution of the sensor. But the lens should always be fitted to match the sensor. This is especially critical for cameras with 5 megapixels or more. Cameras with 10 megapixels or more should use lenses made for professional photography, but those are very expensive.

## 3.8 BEST PRACTICES

---

In rapidly growing markets, new products are introduced by new vendors at a very high frequency. The network video market is no exception. There are currently more than hundreds of different network camera brands. Because network cameras include much more functionality than analog cameras, choosing the right camera becomes not just more important but also more difficult.

The considerations to make include the following:

- *Camera type*: Should it be a fixed, fixed dome, or PTZ camera? Perhaps a panoramic camera is more suitable?
- *Indoor or outdoor environment*: Where will the camera be located? In an outdoor environment, anything but an auto-iris lens is totally impractical and a protective housing is mandatory.
- *Lighting conditions*: What is the lighting in the scene? Is a day-and-night camera necessary? Does it need backup IR lights? Is a low-light camera that provides colors at night needed? Perhaps a thermal camera is more suitable?
- *Resolution*: Higher resolution provides better image detail but comes with some disadvantages. What is the optimal resolution for each application?
- *Intelligence*: Is built-in intelligence such as a tampering alarm, video motion detection, or people counting a requirement?
- *Network functionality*: Does the chosen network camera have all the appropriate networking and security protocols required in today's demanding enterprise networking?
- *Vendor*: What is the brand promise? With hundreds of brands on the market, will the chosen camera brand be on the market years from now? Is the quality high enough? What is the warranty, and can it be extended?



## CHAPTER 4

# Camera technologies

Cameras are complex mechanical and electrical devices that are based on many different technologies. To be able to make the most of a video surveillance system, it is crucial to understand camera technologies.

One of the most important factors—some would argue the most important—of any camera is image quality. This especially is true in video surveillance contexts, where lives and property may be at stake. But what is the definition of image quality, and how can image quality be measured and guaranteed?

There are at least three definitions of image quality:

1. How aesthetically pleasing is the image to the eye?
2. How well does the image reflect reality?
3. How well does the image meet the purpose? For example, the purpose may be to capture a good overview of occurring events or to enable identification of a person or object.

Although one image might not be able to meet all three criteria, most would agree that in a surveillance context, it is important that the image is clear, sharp, and correctly exposed and that it actually delivers critical information. In fact, in a surveillance context, how pleasing an image is to the eye is of little importance. In video surveillance, you need contrast (the ability to discern contrast between features, background, and subject) and sharpness even when there is a lot of movement or when the lighting is poor.

How well a network camera performs depends on many factors. To be able to predict and judge image delivery, one must understand the process and the elements that influence image generation.

This chapter provides a discussion about light, lens, and image sensors and their relation to network cameras, scanning techniques (deinterlacing), image processing (including wide dynamic range [WDR]), and resolution. Chapter 6 discusses compression, which also greatly impacts the quality of recorded and streamed video images.

### 4.1 LIGHT

Light plays a major role in the quality of an image. This section discusses the properties of light and how light affects image quality. It also discusses illuminance and what a camera's lux measurement means. Finally, it explains how cameras can take advantage of near-infrared (NIR) light to produce good-quality, black-and-white images in low-light environments.

### 4.1.1 Light characteristics

Visible light comes in different forms, different directions, and different color hues—all of which affect image quality.

The following are some common forms of light in a scene:

- *Direct light* from a point source or small bright object (such as sunlight or a spotlight) creates sharp contrasts with highlights and shadows.
- *Diffuse light* is light from a source that is so much larger than the subject that it illuminates the subject from several directions (e.g., gray sky, an illuminated screen, a diffuser, or light bouncing off a ceiling). Diffuse light lowers contrasts, which affects the brightness of colors and the level of detail that can be captured.
- *Specular reflection* is light from one direction that bounces off a smooth surface and then is reflected in another direction (such as reflection off water, glass, or metal). Specular reflections within an image can present problems and reduce visibility. A polarizing filter in front of a camera lens can sometimes reduce such reflections.

The directions of the light sources in relation to the subject are also vital. Light direction is a factor when determining how much detail can be obtained from an image. The following are the main light directions:

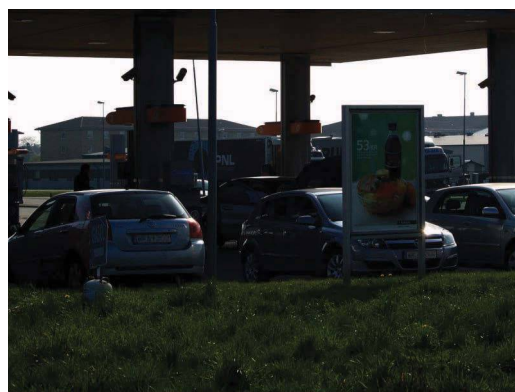
- *Frontal light*: Light that comes from behind the camera. The scene in front of the camera is well illuminated. This is the ideal lighting situation.
- *Sidelight*: Light that illuminates the scene from the side. This may create great architectural effects but also produces shadows.
- *Backlight*: Light comes from or from behind the scene and straight into the camera lens. This light direction is difficult to handle. Silhouettes of objects can be created, and details and color can be lost.

To manage difficult light situations, try to avoid backlight or add artificial light sources. In indoor installations, where backlight or reflection is usually present (e.g., light from large windows®), add frontal lighting. Use diffusers or reflectors to create good illumination. For a discussion on how cameras handle backlight situations and scenes with complex or high-contrast lighting conditions, see Section 4.5.

Figure 4.1a through g shows images that illustrate the effects of different light directions.



(a)



(b)

**Figure 4.1** *Frontal sunlight*: details and colors emerge (a). *Backlight*: details and colors are lost as the camera is placed on the opposite side of the gas station (b). (Continued)



(c)



(d)



(e)



(f)



(g)

**Figure 4.1 (Continued)** *Problem A:* windows® can create reflections and bad exposure (c). *Blinds might help:* reflections are gone but image is underexposed (d). *Solution to problem A:* additional frontal lighting—details appear with added frontal light and correct exposure (e). *Problem B:* partial backlight—nonidentifiable text on vending machine (f). *Solution to problem B:* additional frontal lighting (text on vending machine appears with added frontal light) (g).

4.1.2 Illuminance

As a rule, the more light on the subject, the better the image. With too little light, focusing is difficult and the image will be noisy or dark. But with too much light, the image will be overexposed. Also, different forms and different directions of light affect image quality in different ways.

How much illuminance is required to produce a good-quality image depends on the camera and how sensitive to light it is. In other words, the darker the scene, the more sensitive to light the camera has to be. Light sensitivity is often specified in lux (lx). Lux is a measure of illuminance and the specified value corresponds to a level of illuminance in which a camera produces an acceptable image. The lower the lux specification, the more light sensitive the camera.

Normally, at least 200 lux is needed to illuminate an object so that the camera can produce a good-quality image. Although the specification of high-quality camera may promise an image at 0.1 lux, the image quality might not be very high. The 0.1 lux specification only means that the camera can capture an image at 0.1 lux, but it says nothing about the image quality.

4.1.2.1 Definition of lux

Lux (also known as lumen per square meter or meter-candle) is the amount of light falling onto a surface per square meter. In the lux scale, 1 lux is equal to the amount of light falling on a 1 m<sup>2</sup> surface that is 1 m away from a candle. Correspondingly, 10 lux is the amount of light measured at a distance of 1 m from 10 candles. Foot-candle is another unit for illuminance. One foot-candle is equal to 10.7 lux.

Different light conditions offer different illuminance. Surfaces in direct sunlight receive 100,000 lux, whereas surfaces in full moonlight receive 0.1 lux. Many natural scenes have fairly complex illumination, with both shadows and highlights that give different lux readings in different parts of a scene. The light also shifts in both intensity and direction during the day. Therefore, it is important to keep in mind that one lux reading cannot indicate the light condition for a scene as a whole. And it says nothing about the direction of the light.

Illuminance is measured using a lux meter (see Figure 4.3), and a lux measurement is always specified for a particular surface (Table 4.1).

Figure 4.2a through d shows samples of environments with a lux reading for a specific area of a scene.

Always remember that lux meters and cameras do not collect the same light information. So when we talk about illuminance or lux, we refer to how an object in a scene is lit (incident light), not how the light is collected by the camera. A lux meter (see Figure 4.3) measures only visible light and

**Table 4.1** Examples of various levels of illuminance

Illuminance (lux)	Example
0.00005	Starlight
0.0001	Moonless overcast night sky
0.01	Quarter moon
0.1	Full moon on a clear night
10	Candle at a distance of 30 cm (1 ft)
50	Family living room
150	Office
400	Sunrise or sunset
1,000	Shopping mall
4,000	Sunlight at morning
32,000	Sunlight at midday (min)
100,000	Sunlight at midday (max)





(a)



(b)



(c)



(d)

**Figure 4.2** Trees and telephone lines where the foreground (at ground level) is illuminated by 5 lux (a). An office corridor where the floor is illuminated by about 150 lux (b). A shopping mall where the floor is illuminated by about 500 lux (c). A building on a sunny morning where the building is illuminated by about 4000 lux (d).

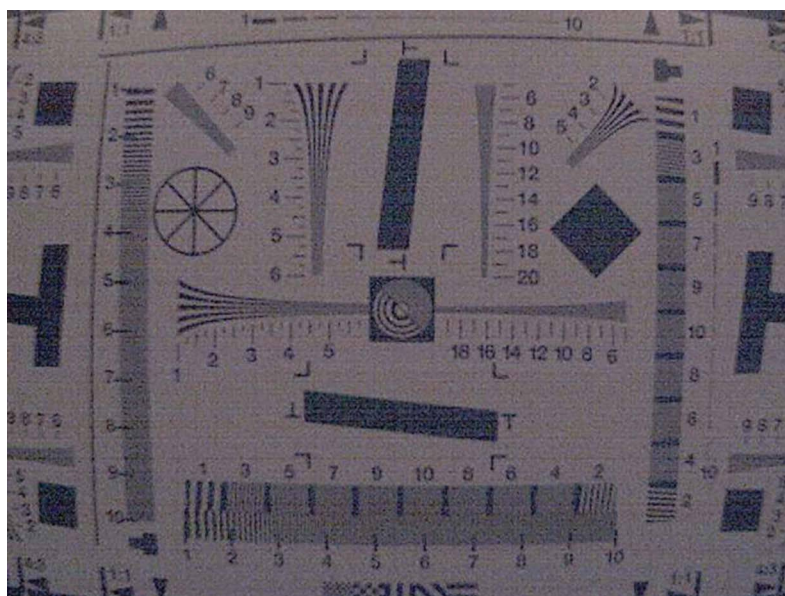


**Figure 4.3** A lux meter is a tool that measures the illuminance of objects and can be used to better understand lighting and its effect on image quality.

it does not take into account the amount of light reflected from an object. For example, the lux readings of two people occupying the same space where one is dressed in white and the other in black would turn out identical, although the white clothes reflect more light than the black clothes. Because it is the amount of light reflected from objects that a camera records, the actual amount of light being captured by a camera may be lower or higher than a particular lux reading due to the reflectance factor. Also, glossy objects reflect more light than dull objects and weather conditions affect lighting and reflection as well. While snow intensifies the reflected light, rain absorbs much of the reflected light.

### 4.1.2.2 Lux rating of network cameras

Many manufacturers specify the minimum level of illumination their cameras need to produce acceptable images. While these specifications are helpful in making light-sensitivity comparisons between cameras produced by the same manufacturer, they are not as useful when comparing cameras from different manufacturers (Figure 4.4). This is because different manufacturers use different methods and have different criteria for what is an acceptable image. The light-sensitivity measurements of day-and-night cameras make matters even more complex. Day-and-night cameras usually have very low lux values because of their sensitivity to NIR light. But because lux is only defined for visible light, it is not totally correct to use lux to express IR sensitivity. Yet doing so is common practice.

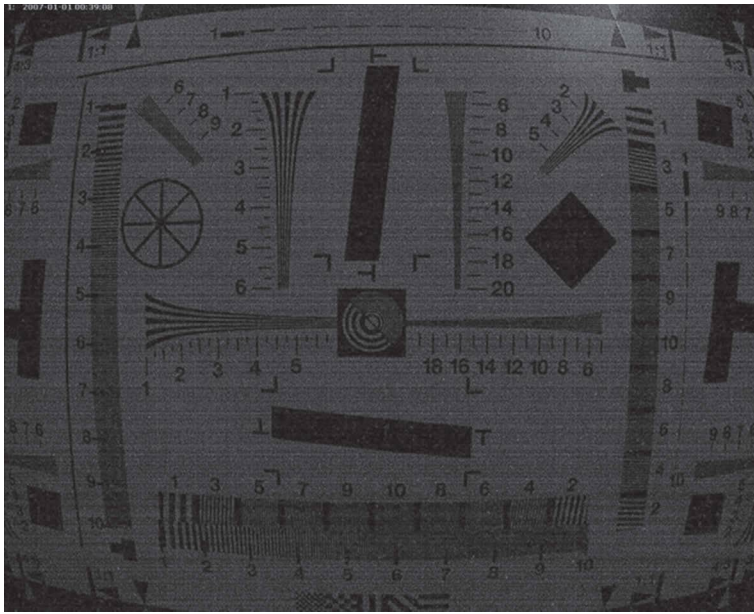


(a)



(b)

**Figure 4.4** The images are from three similarly priced network cameras from different brand-name vendors. The images (a–c) were captured at 1 lux using the cameras' default settings in day mode. Cameras (a) and (b) are specified to produce an image at 1 lux, while camera (c) is specified to produce an image at 0.5 lux. So cameras (a) and (b) are supposed to have the same light sensitivity, and camera (c) is supposed to be the most sensitive to light. Still, at 1 lux, the image from camera (a) clearly has the most brightness and contrast. *(Continued)*



(c)

**Figure 4.4 (Continued)** The images are from three similarly priced network cameras from different brand-name vendors. The images (a–c) were captured at 1 lux using the cameras' default settings in day mode. Cameras (a) and (b) are specified to produce an image at 1 lux, while camera (c) is specified to produce an image at 0.5 lux. So cameras (a) and (b) are supposed to have the same light sensitivity, and camera (c) is supposed to be the most sensitive to light. Still, at 1 lux, the image from camera A clearly has the most brightness and contrast.

#### 4.1.2.3 Lux rating of analog versus network cameras

There are a few differences to consider when measuring or comparing the low-light performance analog and network cameras. The primary difference is the output signal, which in analog cameras is a voltage and in network cameras a digital number. There are differences in how the signal from the sensor is processed as well as some differences in terminology.

Analog video signals are often described in terms of IREs, which is a unit defined by the Institute of Radio Engineers. The IRE values range between 0 and 100. That span corresponds to an approximate voltage difference of 0.7 V in the video signal, or 1.4 V peak to peak. There are some minor differences between the National Television System Committee (NTSC) and Phase Alternating Line (PAL) video systems. For example, in NTSC, 7.5 IRE is defined as black, whereas 0 IRE is used for black in PAL. When the low-light performance is defined as “0.8 lux at 30% IRE” in a datasheet, it means that at 0.8 lux only 30% of the analog voltage can be detected on the BNC output of the analog camera (i.e., 0.21 V, or 0.42 V peak to peak). Because IRE is an analog value, it is irrelevant to network cameras. Still, it is sometimes appended to the digital world through interpretation.

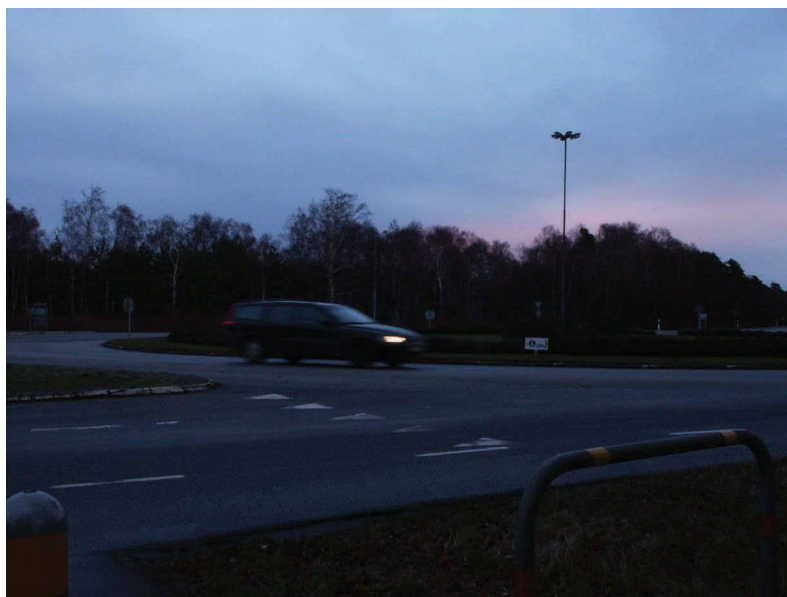
Digital signals consist of digital units, binary numbers. In an 8-bit video, the 7.5–100 IRE analog range is often divided into 256 steps where 0 equals total black (empty pixel) and 255 equals total white (full pixel). However, there are several specifications with their own approaches to what the proper value for black should be and that accommodate the variations in analog video in various ways. Therefore, video software applications use different methods and ranges when digitizing video.

In both analog and network cameras the signal increases with increasing exposure to light. Light exposure can be achieved by increasing the exposure time or by opening the iris. If a brighter image is still needed even though the exposure time and the iris are both at their maximum settings, the signal must be amplified by increasing the gain. Gain does not affect the camera's actual light sensitivity. It is a video signal amplifier. Increasing the gain level boosts the video signal, which



makes the image brighter. The trade-off is that noise increases as well. It is similar to turning up the volume on a radio with poor reception. Not only would the music and speech be louder, there would also be more static and interference (Figure 4.5).

Whenever details, such as the color of hair, eyes, clothing, and vehicles, are necessary to identify subjects and objects, color images have an undeniable advantage over black-and-white images. This is one of the reasons why a camera's light sensitivity is so important. Today, there are sensors that are exceptionally sensitive to light and network cameras with advanced built-in image processing that can filter out the noise that comes from high gain. A camera that combines these two qualities



(a)



(b)

**Figure 4.5** A scene taken at 500 lux with low gain (a). A similar scene also taken at 500 lux but with high gain (b). The images illustrate how increasing the gain amplifies video signal and makes the image brighter at the sacrifice of increased noise in the image.



can provide color video in low-light conditions that would normally force the camera to switch to black-and-white video.

To properly compare the low-light performance of two different cameras, it is necessary to look beyond lux and IRE numbers. The best way is to put the cameras side by side, record the same scene in the same conditions, and compare the outcome.

### 4.1.3 Color temperature

Another thing to consider is how different types of light affect the color of images. Many types of light, such as sunlight and incandescent lamps, can be described in terms of their color temperature (Figure 4.6). The color temperature is measured in Kelvin (K), and the scale is based on the radiation of all heated objects. The first visible light radiating from a heated object is red. As the temperature of the heated object rises, the radiating color becomes bluer. Red has a lower color temperature than blue. Ironically, this is just the opposite of what is sometimes meant with the colors blue and red when indicating hot or cold water. Near dawn, sunlight has a low color temperature (implying redder colors), whereas during the day, it has a higher one (implying yellower, more neutral colors).

At midday, sunlight has a color temperature of about 5500 K, which also is about the temperature of the solar surface. A tungsten light bulb has a color temperature of about 3000 K. Some light sources, such as fluorescent lamps that are gas-discharge lamps rather than heated filaments, are further from the approximation of a radiating heated object. Such light sources cannot be described as accurately in terms of degrees Kelvin. Instead, the closest color temperature is used.

Light-emitting diode (LED) illuminators are high-efficiency lighting sources with advanced control methods for extremely low energy consumption. LEDs are popular complements to network cameras and are used to provide required lighting in demanding scenes. LEDs can be operated using Power over Ethernet, their light can be easily directed using optics, and they have a very long life. The color temperature for white-light LEDs is similar to that of daylight.

Scenes illuminated by light sources with different colors (or spectral distribution) look different to a camera. The human visual system, however, has a remarkable way of coping with such changes so that colored objects appear to maintain their color. This is sometimes referred to as color constancy. For a camera to do the same thing, it must adapt to the local illumination. This process is sometimes referred to as white balancing. In its simplest form, it uses a known object (usually gray) and makes color adjustments to an image so that a gray color (and all other colors) in a scene appears as the human visual system perceives it. Figure 4.7 shows scenes illuminated by different light sources.

Most modern cameras have an automatic white balance system. White balance often can be adjusted manually or selected among presets.

### 4.1.4 Invisible light

As discussed previously, the color (or spectral distribution) changes when the temperature of the light source changes. The colors are still visible in the range from just below 3,000–10,000 K, so we can see them. For cooler or hotter objects, the bulk of the radiation is generated within the invisible wavelength bands.

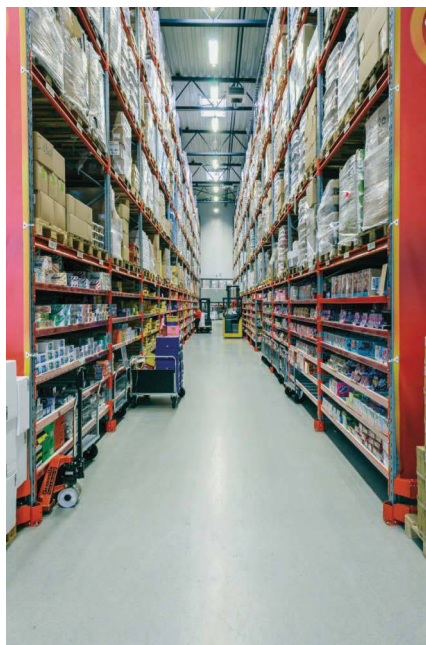
Outside the visible range of light, we find infrared (IR) and ultraviolet (UV) light, which cannot be seen or detected by the human eye. However, most camera sensors can detect some of the NIR



**Figure 4.6** Color temperature of visible light (from highest to lowest).



(a)



(b)



(c)

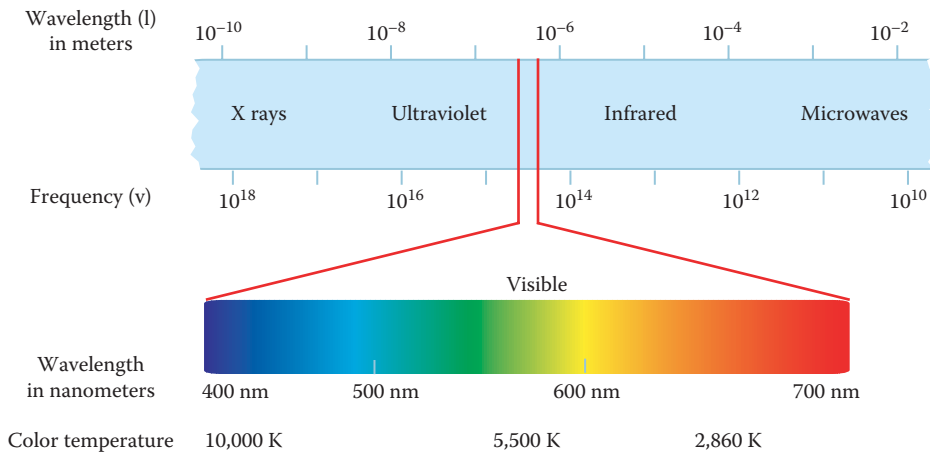


(d)

**Figure 4.7** Scenes illuminated by different light sources: An office common area with indoor office lights. Standard light bulbs (3000 K) are more reddish than daylight and create a brown or yellow tone in the image. To compensate, the camera uses white balancing techniques (a). Warehouse industrial long-tube fluorescent lights are designed to offer unobtrusive light. However, the images will appear with a green and dull tone. Images need white balance compensation (b). Some artificial electrical light, like in a shopping mall, may offer something similar to daylight (c). Natural light has different colors depending on the time of day. Here, a city waterfront at midday (d).

light, from 700 nm up to about 1000 nm. If such light is not filtered out, it can distort the color of resulting images (Figure 4.8).

Therefore, a color camera is outfitted with a filter, which is an optical piece of glass placed between the lens and the image sensor. This IR blocking is commonly called an IR cut filter, and it filters out NIR light and delivers the same color interpretations that the human eye produces. UV light, on the other hand, does not affect a surveillance camera because charge-coupled device (CCD) and complementary metal-oxide semiconductor (CMOS) sensors are not sensitive to this light. However, an analog film camera is sensitive to UV (UVA) light and requires a UV filter or coating on the lens of the camera.



**Figure 4.8** The wavelengths of light. Near-infrared light spans the range between 700 and 1400 nm.

The IR cut filter can be removed to extend a network camera's ability to produce quality images in low-light or dark situations. This allows a camera's image sensor to use the NIR light to deliver high-quality black-and-white images. Cameras with the ability to make use of NIR light often are marketed as day-and-night cameras or IR-sensitive cameras. This does not mean that such cameras produce heat-sensitive IR images. IR images require true IR cameras that are specialized at detecting long-wave infrared (LWIR) light (heat) that radiates from both living and nonliving objects. In IR images, warmer objects, such as people and animals, stand out from typically cooler backgrounds. True IR cameras are called thermal cameras. For more information about thermal cameras, see Chapter 5. For more information about day-and-night cameras, see Chapter 3.

## 4.2 LENSES

The first camera component to capture the light is the lens, and therefore it has a big impact on image quality. A camera lens normally consists of an assembly of lenses. The camera lens functions that affect image quality include:

- *Defining the field of view*, that is, defining how much of a scene and the level of detail that the camera shall capture
- *Controlling the amount of light* that passes through to the image sensor so that the image is correctly exposed
- *Focusing* by adjusting either elements within the lens assembly or the distance between the lens assembly and the image sensor

This section discusses different types of lenses, as well as lens characteristics such as field of view, iris, f-number, focusing, mounts, and lens quality.

### 4.2.1 Lens types

There are three main types of lenses:

1. Fixed lens
2. Varifocal lens
3. Zoom lens

A fixed lens (see Figure 4.9) has only one field of view because the focal length is fixed. There is a choice between a normal, telephoto, or wide-angle view. Therefore, to choose the correct lens, one has to know beforehand exactly which focal length is needed. A 4-mm focal length is common among fixed network camera lenses.



**Figure 4.9** Fixed lens.



**Figure 4.10** Varifocal lens.

A varifocal lens (see Figure 4.10) offers a range of focal lengths and therefore different fields of view. The field of view can be adjusted. Whenever changing the field of view, the user must also manually refocus the lens. Varifocal lenses for network cameras often have focal lengths that range from 3.0 to 8 mm.

Zoom lenses, like varifocal lenses, enable the user to select different fields of view. But a zoom lens has autofocus, which means that there is no need to refocus the lens if one changes the field of view. Focus is maintained within a specified focal length range, for example, 6–48 mm. Lens adjustments can be either manual or motorized and remote controlled. The zoom capacity of a lens refers to the ratio between the lens' longest and shortest focal lengths. For example, a lens with three times zoom has a maximum focal length that is three times longer than its shortest focal length.

#### 4.2.1.1 IR-coated lenses

IR-coated lenses are treated with special materials to provide IR correction for color and IR light. The coating also has antireflective qualities that improve the properties of the IR light that passes through the lens and reaches the sensor.

### 4.2.2 Lens mount standards

There are three main mount standards used for interchangeable network camera lenses: CS-mount, C-mount, and S-mount.

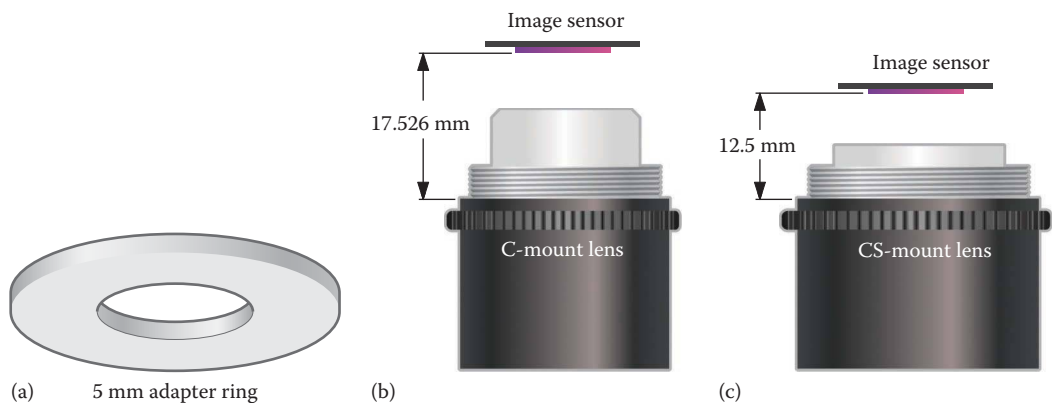
CS-mount and C-mount look the same and both have a 1-inch thread and a pitch of 32 lines per inch. The screw mount size is defined as 1–32 UN 2A by the unified thread standard established by the American National Standards Institute. CS-mount is an update to the C-mount standard that allowed for reduced manufacturing cost and sensor size. Today, CS-mount is much more common than C-mount. Sometimes, manufacturers make the thread of their cameras' lens mount slightly smaller or larger than the standard to ensure perfect alignment between the lens and the sensor.

What differs CS-mount from C-mount is the flange focal distance (FFD), which is the distance from the mounting flange to the sensor when the lens is fitted on the camera.

- *CS-mount*: the FFD is 12.5 mm ( $\approx 0.49$  in.  $\approx 1/2$  in.).
- *C-mount*: the FFD is 17.526 mm (0.69 in.  $\approx 11/16$  in.).

If it is impossible to focus the camera, one probably has the wrong type of lens. A C/CS adapter ring, which in essence is a 5 mm spacer, can be used to convert a C-mount lens to a CS-mount lens (Figure 4.11).

S-mount (Figure 4.12) is also called M12-mount and is common in small cameras such as covert cameras and fixed mini-dome cameras. It is called M12 because it has a nominal outer thread diameter of 12 and 0.5 mm pitch (M12  $\times$  0.5). Thus, it follows the metric screw thread standard established by the International Organization for Standardization. Because S-mount lenses are often mounted directly on the circuit board, they are sometimes called board lenses.



**Figure 4.11** A C/CS adapter ring (a), a C-mount lens (b), and a CS-mount lens (c).



**Figure 4.12** S-mount lens.



Thermal cameras often use other lens mounts than C- or CS-mount. The first documented thermal lens standard is TA-lens. The letters “TA” stand for Thermal A, where “A” stands for the first documented standard. The lens mount is an  $M34 \times 0.5$  screw mount. It is large enough for sensors with a diameter up to at least 13 mm, which means that it works for many popular LWIR sensors.

Fixed dome cameras often use a type of board lens that has a round high-precision surface with a diameter of 14 or 19 mm. Aligning this type of lens is a challenge and therefore it is best mounted with a custom-made holder. These lenses are not interchangeable.

Pan, tilt, and zoom (PTZ) cameras use modules—known as camera blocks or block lenses—that consist of a sensor and lens assembly with an auto-iris, as well as a number of mechanical parts for autofocusing and motorized zooming. PTZ camera lenses are not interchangeable.

### 4.2.3 Field of view (focal length)

It is important to know what field of view (FoV) the camera must cover because it determines the amount and level of information that can be captured in an image. Field of view, also called angle of view, is divided into several subareas where the most common in network video are the following (also see Figure 4.13):

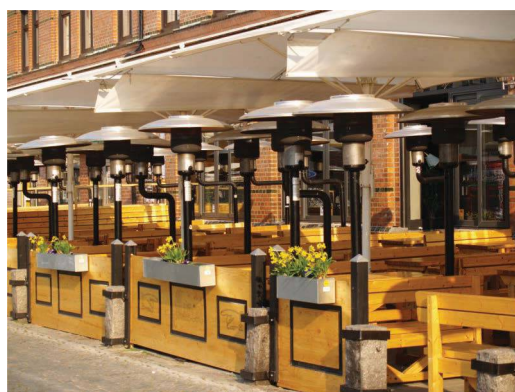
- *Normal view* means that the image has the same field of view as the human eye.
- *Telephoto view* is a magnification of a narrower field of view, generally providing finer details than a human eye can deliver.
- *Wide-angle view* is a larger field of view with fewer details than a normal view.



(a)



(b)



(c)

**Figure 4.13** Different fields of view: wide-angle view (a), normal view (b), and telephoto—narrow field of view (c).

The focal length of the lens and the size of its image sensor determine a network camera's field of view. Focal length is defined as the distance between the entrance lens (or a specific point in a complex lens assembly) and the point where all the light rays converge to a point, typically the camera's image sensor. The longer the focal length, the narrower the field of view. To achieve a wide field of view, the focal length should be shorter.

To understand terms such as field of view, focal lengths, and normal, telephoto, and wide-angle lenses, it helps to compare with a traditional camera that uses 35-mm film. The human eye has a fixed focal length that is equivalent to a lens with a focal length of 50 mm on a classic 35-mm film camera. Traditional lenses with focal lengths ranging from 35 to 70 mm are therefore considered "normal" or "standard" lenses that have a normal field of view (Figure 4.14a).

Telephoto lenses (Figure 4.14b) for traditional cameras have focal lengths of more than 70 mm. A telephoto lens is used when the surveillance object is either small or located far away from the camera. A telephoto lens delivers good detail for long-distance viewing. Usually, telephoto images have a slight geometrical distortion, which appears as curvatures at the edges. Generally, a telephoto



**Figure 4.14** Network camera lenses: (a) normal lens with standard focal length, (b) telephoto lens with long focal length, and (c) wide-angle lens with short focal length.

lens has less light-gathering capability and therefore requires that the scene has good lighting to produce a good-quality image.

Wide-angle lenses (Figure 4.13c) have focal lengths of less than 35 mm. The advantages include a wide field of view, good depth of field, and decent low-light performance. The downside is that this type of lens produces geometrical distortions. Lenses with a focal length of less than 20 mm create what is often called a fish-eye effect. Wide-angle lenses are not often used for long-distance viewing.

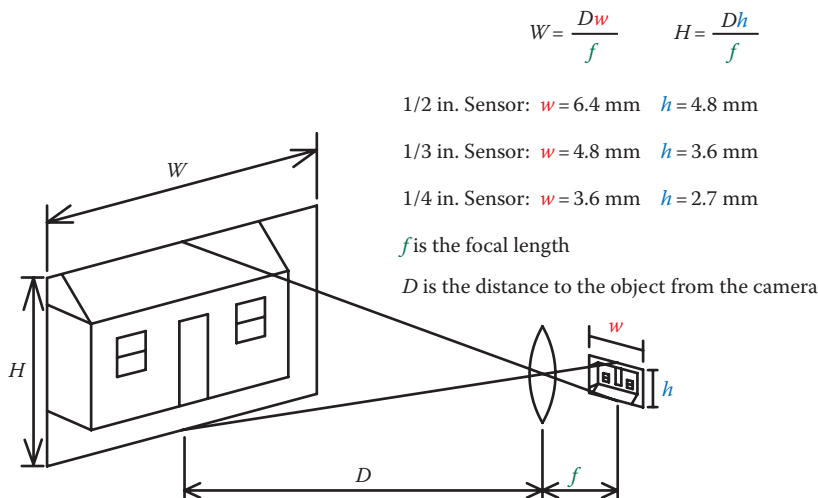
A network camera’s field of view is determined by both the lens’ focal length and the size of its image sensor. Therefore, the focal length of a network camera lens is irrelevant unless one also knows the size of the camera’s image sensor, which can vary. The focal length of a network camera lens is generally shorter than the focal length of its counterpart among classic photographic camera lenses. This is because the size of an image sensor (the digital equivalent to a 35-mm film) is much smaller than the size of a frame on a 35-mm film.

The typical sizes of image sensors are ¼, ⅓, ½, and ⅔ in. There are two ways to determine the field of view for a given network camera lens. The easy way is to put it in relation to a traditional 35-mm camera. The size of the image sensor has a given conversion factor. Multiplying the conversion factor with the focal length of the network camera lens gives what the focal length of an equivalent 35 mm camera lens would be (see Table 4.2).

The second way is to calculate it using the formula in Figure 4.15. Given the focal length of the lens and the distance to the scene or subject from the camera position, this formula helps determine the height and width of the scene that will be captured.

**Table 4.2** Field of view for a given network camera

Sensor = factor × network camera lens = traditional camera’s focal length			
in.	in.	mm	mm
1/4	9	3–8	27–72
1/3	7	3–8	21–56
1/2	5	3–8	15–40



**Figure 4.15** In the formula, (w) and (h) correspond to the physical width and height of the image sensor.



### Calculation in feet:

When using a camera with a  $\frac{1}{4}$  in. CCD sensor and a 4-mm lens ( $f$ ), how wide ( $W$ ) does an object have to be to be visible at 10 ft ( $D$ )? A  $\frac{1}{4}$  in. sensor is 3.6 mm wide ( $w$ ).

### Calculation in meters:

When using a camera with a  $\frac{1}{4}$  in. CCD sensor and a 4-mm lens ( $f$ ), how wide ( $W$ ) does an object have to be to be visible at 3 m ( $D$ )? A  $\frac{1}{4}$  in. sensor is 3.6 mm wide ( $w$ ).

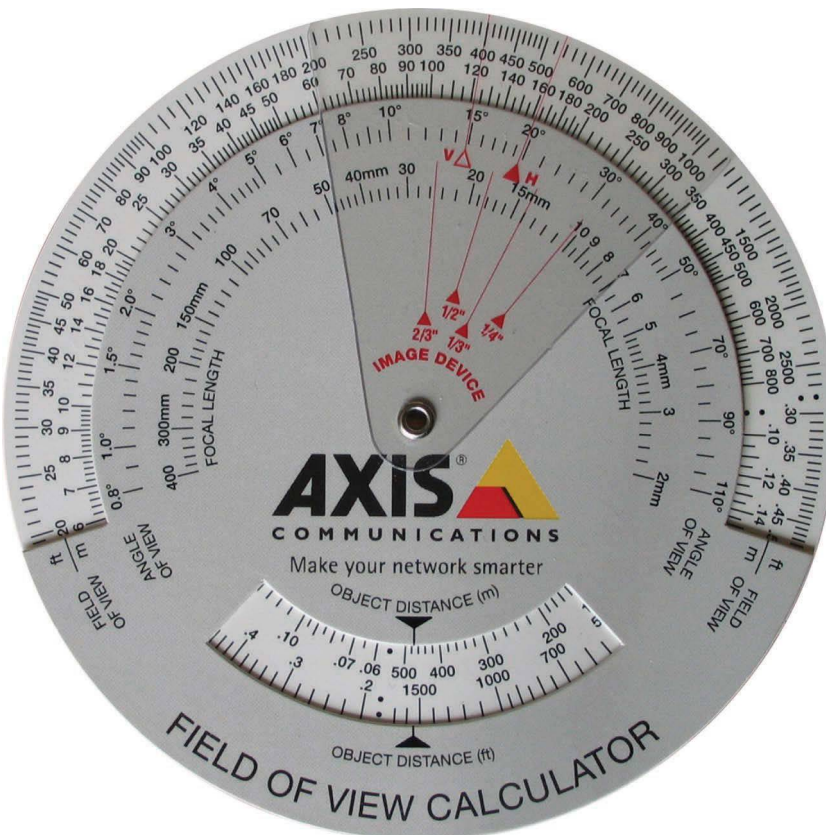
$$W = D \times \frac{w}{f} = 3 \text{ m} \times \frac{3.6 \text{ mm}}{4 \text{ mm}} = 2.7 \text{ m}$$

The fastest way to calculate the field of view and to determine which lens is required is to use an online lens calculator (Figure 4.17) or a rotating lens calculator (Figure 4.16).

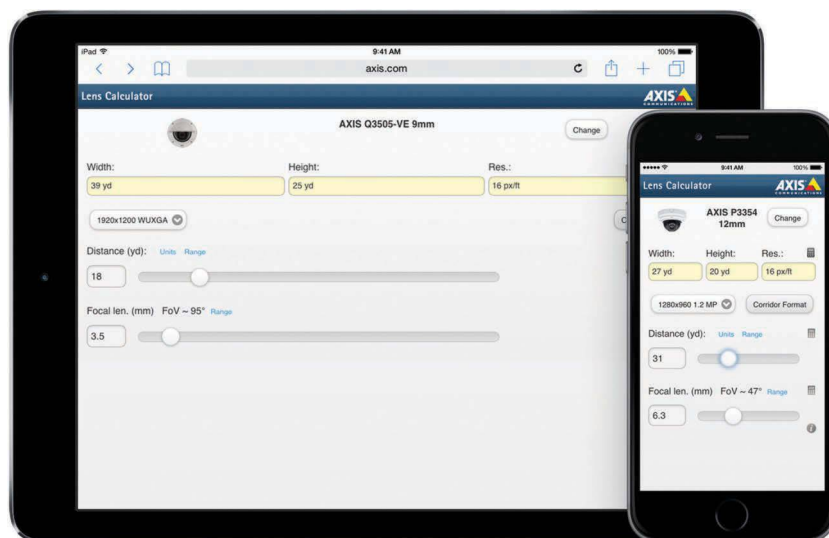
Several manufacturers make their lens calculators available online. These tools make calculations quick and convenient (Figure 4.17). The drawback is that they ignore the potential geometrical distortions of lenses.

## 4.2.4 Matching lens and sensor

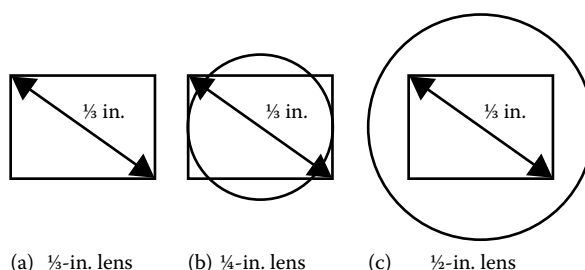
As mentioned earlier, image sensors are available in different sizes, such as  $\frac{2}{3}$ ,  $\frac{1}{2}$ ,  $\frac{1}{3}$ , and  $\frac{1}{4}$  in. Lenses are manufactured to match these sizes. Selecting a lens that is suitable for the camera is fundamental (Figure 4.18). A lens made for a  $\frac{1}{2}$ -in. sensor works with  $\frac{1}{2}$ -,  $\frac{1}{3}$ -, and  $\frac{1}{4}$ -in. sensors, but not with a  $\frac{2}{3}$ -in. sensor.



**Figure 4.16** A rotating lens calculator is a good tool for quickly calculating the lens required or the field of view.



**Figure 4.17** With a tablet or smartphone that is connected to the internet, an online lens calculator can be accessed from any location. Here is an example of an online calculator viewed in a browser on iPad® and iPhone®.



**Figure 4.18** Examples of different lenses (a through c) paired with a  $\frac{1}{3}$ -in. sensor.

If using a lens that is made for a smaller sensor than the sensor that is actually fitted inside the camera, the image will have black corners. If using a lens that is made for a larger sensor than the sensor that is actually fitted inside the camera, some information will be lost because it falls outside of the sensor. Therefore, the field of view will be smaller than what the lens allows for. The image will appear to have a telephoto effect, making everything look zoomed in. This is known as a focal length multiplier greater than 1. Assuming that the resolution is the same, the smaller the image sensor for a given lens, the more a scene is magnified (see Figure 4.19).

#### 4.2.5 Aperture (iris diameter)

A lens' ability to let light pass through is measured by its aperture or iris diameter. The bigger the aperture of a lens, the more light can pass through it. In low-light environments, it is generally better to have a lens with a large aperture.

To get a good-quality image, the amount of light passing through a lens must be optimized. If too little light passes through, the image will be dark. If too much light passes through, the image will be overexposed.

The amount of light captured by an image sensor is controlled by two elements that work together:

1. *Aperture*: The size of the lens' iris opening.
2. *Exposure time*: The length of time an image sensor is exposed to light.



(a)

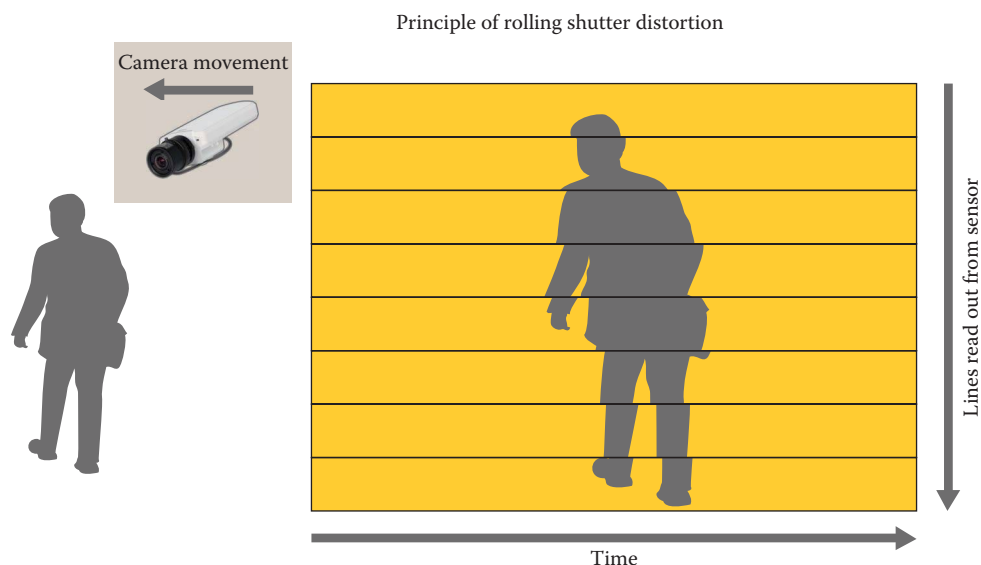


(b)

**Figure 4.19** Field of view for same lens and resolution, but different sensor size: Image (a) is taken using a  $\frac{1}{3}$ -in. sensor. Image (b) is taken using a  $\frac{1}{4}$ -in. sensor. Because the sensor of image (b) is smaller than  $\frac{1}{3}$  in., the image covers a smaller part of the scene. It corresponds to the parts of image (a) that are contained within the red border. In other words, the result of using a lens that is too big for the sensor is a magnification of a smaller area.

The longer the exposure time, the more light an image sensor receives. Well-lighted scenes need shorter exposure time, whereas darker scenes need longer exposure times. A certain exposure level can be achieved using either a large iris opening and a short exposure time or a small iris opening and a long exposure time. It is important to know that increasing the exposure time also increases motion blur and increasing the iris opening reduces the depth of field.

As mentioned before, gain is another parameter that affects how much light the sensor is exposed to. Gain is a signal amplifier; that is, it tunes the sensitivity of the image sensor. But increasing the gain increases the noise level.



**Figure 4.20** The principle of rolling shutter distortion. A shaking camera causes misalignment between the scanned lines.

In traditional cameras, the shutter controls the exposure time. However, most network cameras do not use mechanical shutters. Instead, the image sensor acts as an electronic shutter. It switches on and off to collect and discharge electrical charges. There are two types of electronic shutters. Many network video cameras, especially cameras with CMOS sensors, come with a rolling shutter. Unlike a global shutter, which exposes all pixels at the same time in a single snapshot, the rolling shutter catches the image by scanning across the frame, line by line. In other words, not all parts of the image are captured at the same time, but each line is exposed during a slightly different time window. Therefore, if the camera shakes or vibrates, each exposed line is slightly moved in relation to the other lines, which makes the image warp or wobble. With rolling shutters, fast-moving subjects may also appear distorted (see Figure 4.20).

Rolling shutter distortion induced by vibrations can be avoided with stabilization techniques. Optic stabilization instantaneously compensates for the motion. With electronic stabilization, the rolling shutter must first scan at least one line before the digital processing to stabilize the image can begin. Even so, this method works very well and the technology is improving rapidly. For more information about image stabilization, see Section 3.3.1.

#### 4.2.6 Types of iris control

The ability to control a camera's iris opening plays a central role in image quality. An iris is used to maintain the optimum light level to the image sensor so that images are properly exposed. The iris can also be used to control the depth of field, which is explained in more detail in Section 4.2.8. Iris control can be fixed or adjustable, and adjustable iris lenses can be manual or automatic. Automatic iris lenses can further be classified as either auto-iris or P-Iris lenses.

With fixed iris lenses, the iris opening cannot be adjusted and is fixed at a certain f-number. The camera can compensate for changes in light level by adjusting the exposure time or using gain.

With manual iris lenses, the iris can be adjusted by turning a ring on the lens to open or close the iris. This makes them less suitable for outdoor surveillance or other environments with changing light conditions.

There are two types of auto-iris lenses: DC-iris and video iris. Both use a galvanometer to automatically adjust the iris opening in response to changes in light levels. To control the iris opening,



they use an analog signal. They differ in the location of the circuitry that converts the analog signal into control signals. In a DC-iris lens, the circuit resides inside the camera. In a video iris, the circuit is inside the lens.

In bright situations, a camera with an auto-iris lens can be affected by diffraction and blurring when the iris opening becomes too small. This problem is especially prominent in megapixel and high-definition television (HDTV) cameras because the pixels in those image sensors are smaller than the pixels in lower-resolution cameras. Therefore, the image quality is dependent on getting the right aperture. In order to optimize image quality, a camera needs to have control over the iris opening. The problem with an auto-iris lens is that this control cannot be made available to the camera or user.

P-Iris was designed to address the shortcomings of an auto-iris lens while maintaining the benefits of an automatically controlled iris. It involves a P-Iris lens, a motor that allows the position of the iris opening to be precisely controlled, as well as specialized software for optimized image quality.

In bright situations, what typically happens with a camera that uses a DC-iris lens and a megapixel sensor with small pixels is that it produces blurry images because the aperture becomes too small. This type of image blur is called diffraction. P-Iris limits the closing of the iris to prevent diffraction.

P-Iris provides improvements in contrast, clarity, resolution, and depth of field. Having good depth of field (objects at different distances from the camera are in focus simultaneously) is crucial when monitoring areas with a good amount of depth in them, such as long corridors, server halls, or parking lots.

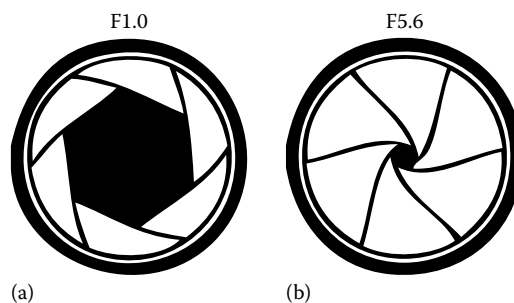
#### 4.2.7 F-number (f-stop)

In low-light situations, particularly in indoor environments, a vital factor to look for in a network camera is the lens' ability to collect light. The *f*-number is a measure of this light-gathering ability. It is the ratio (*N*) of the lens' focal length (*f*) to the diameter of the aperture (*D*):

$$N = \frac{f}{D}$$

F-numbers can be expressed in several ways. The two most common ways are *f*/*D* and *FD*. The slash indicates division as explained in the formula given earlier. An *f*-number of *f*/4, or F4, means that the aperture is equal to the focal length (*f*) divided by 4. So if a camera has an 8 mm lens, light must pass through an aperture that is 2 mm in diameter.

The smaller the *f*-number, the better the lens' light-gathering ability. That is, more light can pass through the lens to the image sensor. A small *f*-number is achieved through a large aperture relative to the focal length (see Figures 4.21 and 4.23).



**Figure 4.21** A smaller *f*-number means a larger aperture (a), while a larger *f*-number means a smaller aperture (b).

**Table 4.3** Amount of light relative to the F5.6 f-stop

F-stop	F1.0	F1.4	F2.0	F2.8	F4.0	F5.6
Relative light level	32	16	8	4	2	1

The difference in light between two sequential f-numbers is called an f-stop. Traditional camera lenses usually use a standardized sequence (F1.0, F1.4, F2.0, F2.8, F4.0, F5.6, F8.0, etc.) where each f-stop represents a halving of the light intensity from the previous stop. This is illustrated in Table 4.3, which shows the amount of light relative to the F5.6 f-stop. This means that an F1.0 lens lets through 32 times more light than an F5.6 lens. In other words, it is 32 times more light sensitive.

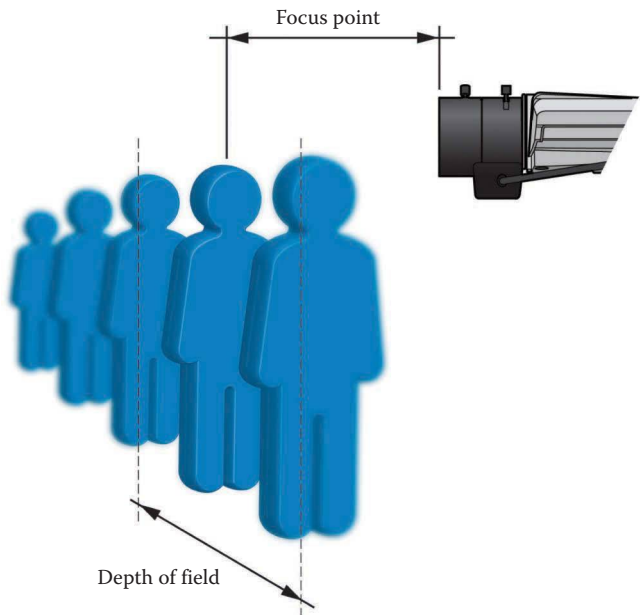
In low-light situations, a smaller f-number generally produces a better image quality. However, some sensors may not be able to take advantage of a lower f-number in low-light situations because of the way they are designed. And choosing a lens with a higher f-number means an increase of the depth of field, which is explained in Section 4.2.8.

Although network cameras with DC-iris lenses have an f-number range, often only the maximum light-gathering end of the range is specified. A lens with a lower f-number is usually also more expensive than a lens with a higher f-number.

**4.2.8 Depth of field**

A lens can only focus at a single point, but there is an area that stretches in front of and behind the focal plane that still appears sharp. This volume of space is known as the depth of field (see Figure 4.22). The depth of field is not a fixed distance or volume of space. It changes in size and is usually described as either short (shallow) or long (deep). In an image with short depth of field, only a small zone appears sharp. In an image with long depth of field, a larger zone appears sharp.

Depth of field is important in many video surveillance situations. For example, when monitoring a parking lot, the camera needs to be able to deliver sharp images of cars and their license plates at a range of distances, say 20, 30, and 50 m (60, 90, 150 ft) away.



**Figure 4.22** Imagine a line of five people. The camera focuses on the second person. The faces of the person behind and the person in front can also be identified. They stand within the depth of field.

**Table 4.4** How focal length, aperture, and camera-to-object distance affect depth of field

	Short depth of field	Long depth of field
Focal length	Long (telephoto lens)	Short (wide-angle lens)
Aperture	Large (small f-number)	Small (big f-number)
Distance to object	Short	Long

Depth of field is affected by a number of factors. The ones that are most often mentioned and that can be controlled by the user are as follows:

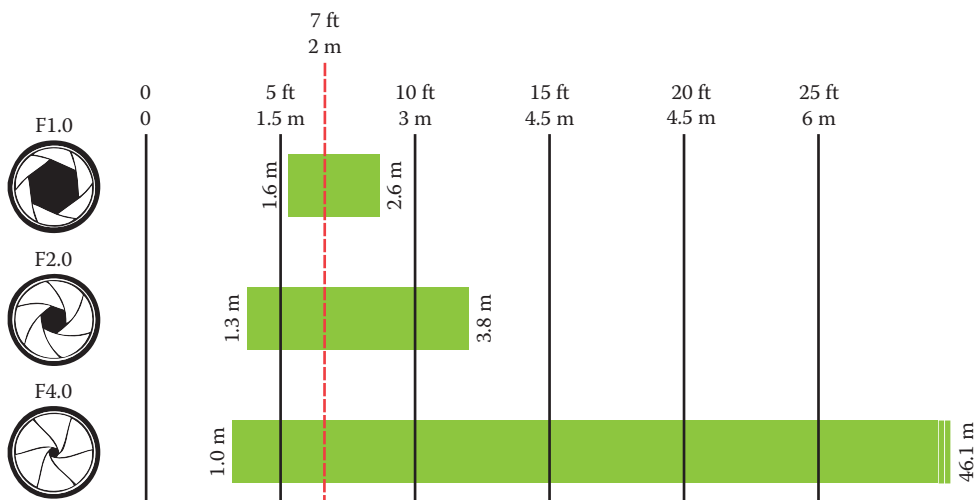
- *Focal length*: The longer the focal length, the shallower the depth of field.
- *Aperture (iris diameter)*: The larger the aperture, the shallower the depth of field.
- *Camera-to-object distance*: The shorter the distance, the shallower the depth of field.

The longer the focal length or the shorter the distance of the camera to the object, the shorter the depth of field. Depth of field becomes shorter as the aperture increases. Thus, the smaller the aperture, the better. Table 4.4 gives a summary of the effect that focal length, aperture, and camera-to-object distance have on depth of field.

Note that unlike aperture and focal length, camera-to-object distance is not set by physical camera parameters, but is simply an effect of how close the camera is positioned in relation to the object. If the camera's position is limited by the surrounding elements, such as poles or buildings, and the range at which objects need to be in focus is nonnegotiable, as is often the case in video surveillance situations, what is left to play with is the lens' focal length and aperture.

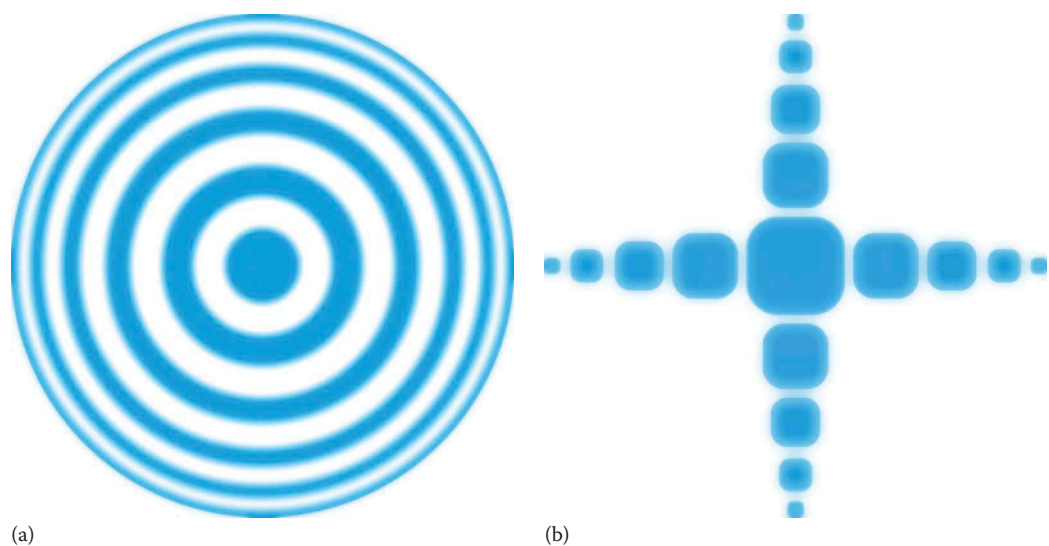
Figure 4.23 gives an example of the depth of field for different f-numbers at a focal distance of 2 m (7 ft). A large f-number (smaller aperture) enables objects to be in focus over a longer range.

The depth of field changes gradually. It does not abruptly shift from sharp to unsharp. Everything that is immediately in front or behind of the focusing plane gradually loses sharpness. To our eyes, the points that are projected within the depth of field still appear sharp (see also Figure 4.22). The circle of confusion is a term that is used to define how much a point needs to be blurred to be perceived as unsharp. When our eyes are able to see the circle of confusion, those points are no longer acceptably sharp and are outside the depth of field.



**Figure 4.23** Iris opening in relation to depth of field. Depending on the pixel size, very small iris openings may blur an image due to diffraction.





**Figure 4.24** Circular apertures that are too small produce a diffraction pattern of concentric rings (a), whereas square apertures produce a cross pattern (b).

In low-light situations, there is a trade-off between achieving good depth of field and reducing blur from moving subjects. To get a good depth of field and a properly exposed image in low light, a small aperture and long exposure time are necessary. However, the long exposure time increases motion blur. A smaller aperture generally increases the sharpness because it is less prone to optical errors. But if the aperture is too small, diffractions can appear in the image. With round apertures, the diffractions produce a pattern of circles that is called an Airy disk or Airy pattern (named for George Biddell Airy). So in an image of a round object, the object would be surrounded by fainter rings of color (Figure 4.24). Similar blurring effects occur with other aperture shapes, but the geometrics of the aperture would render another pattern than a circular one.

How to get the optimal result is a matter of weighing the environment against the image requirements and priorities (see Table 4.5).

The following are some guidelines of what is achievable under various lighting conditions:

- *Sunshine*: With lots of light, the camera chooses a small iris opening and a short exposure time. This gives maximum image quality with good depth of field and clear image without any motion blur.
- *Overcast*: If depth of field is a priority, increase the exposure time to reduce the aperture. But note that blurring of moving subjects increases. For clearer images of moving subjects, reduce the exposure time. But note that depth of field is compromised through a bigger aperture.
- *Evening and night*: In low-light situations, the camera adjusts the gain of the image sensor to deliver good images. But note that noise levels increase and appear as grainy effects in the image.

**Table 4.5** What is achievable under various lighting conditions

	Sunshine	Low-light setting 1	Low-light setting 2
Exposure time	Short	Long	Short
Aperture	Small (big f-number)	Small (big f-number)	Large (small f-number)
Gain or sensitivity	Limited	Limited	Full
Image priority	All	Depth of field	Movement
Result	Good depth of field, clear image, frozen action <sup>a</sup>	Good depth of field, clear image, blurry movement <sup>a</sup>	Limited depth of field, noise, frozen action

<sup>a</sup> A very small iris may decrease sharpness.

### 4.2.9 Focusing

Focusing a network camera often requires making really fine adjustments to the lens. With some auto-iris network cameras, one might encounter problems focusing if attempting to set focus under very bright or very dark conditions.

With less sophisticated auto-iris lenses, it is best to adjust the focus of an auto-iris lens in low-light conditions or by using a darkening filter such as a neutral density filter or even a welder's glass. The filter fools the camera that it is darker than it really is. In low-light conditions the iris automatically opens, which shortens the depth of field and helps the user focus more precisely. The more sophisticated auto-iris lenses used in many modern IP cameras allow you to open the iris through the camera's user interface. In those cases, a darkening filter is not needed.

Given a longer depth of field, it may be easier to set the focus in bright conditions. But as light decreases and the iris diameter increases, the depth of field becomes shorter and the image blurry.

To ensure optimal performance in all situations, the market provides a wide range of lenses. When choosing and adjusting a camera lens, it is best to follow the manufacturer's instructions.

Some cameras come with autofocus. This means that the camera automatically adjusts the lens mechanically so that the image is focused. The autofocus feature is a requirement in pan-tilt cameras where the camera direction is constantly changing. Some fixed cameras also have autofocus, but because focusing a camera is normally only a one-time installation thing, it is often difficult to justify the additional cost. Sometimes these cameras have a focus window, which sets up focus positions that allow for easier autofocusing.

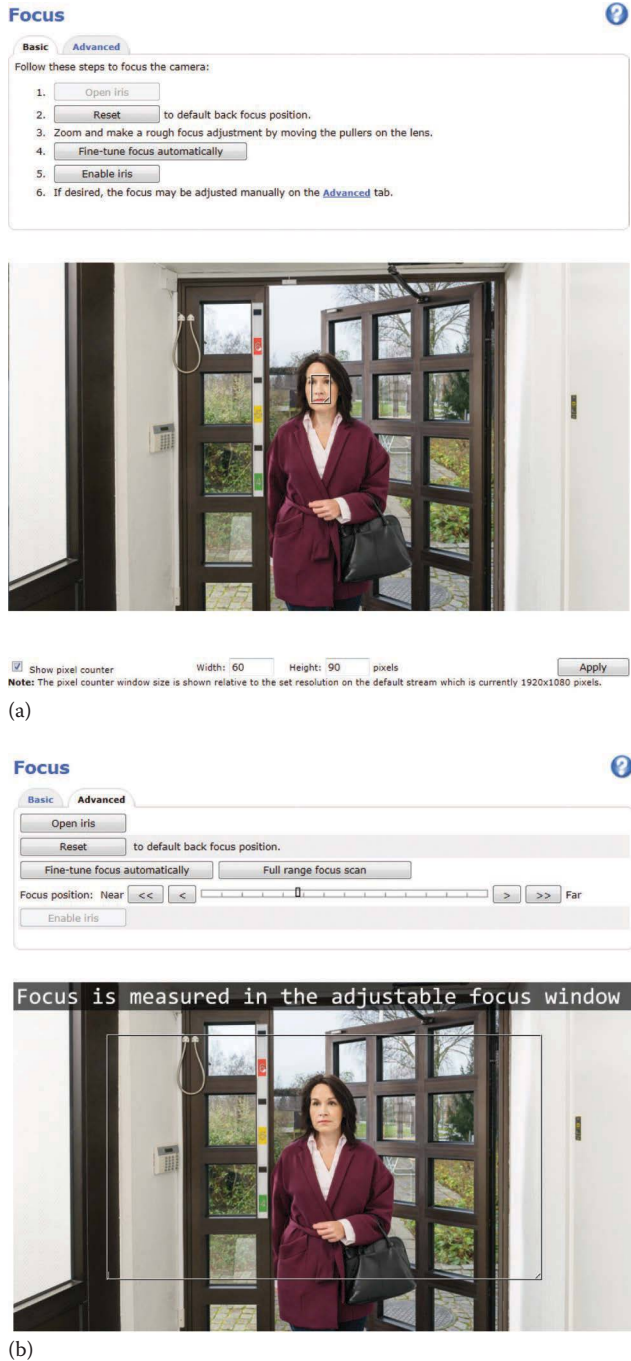
If the camera has remote focus, there is no need to worry about the time-consuming task of manually adjusting the focus on-site. The installer can use the focus puller for a rough focusing in the field (Figure 4.25). Then exact adjustment of the focus can be done remotely through a computer. Some cameras include full remote focus capabilities, eliminating the need for manual adjustment on-site. The computer window gives live feedback so that the user sees that the focus is right. See Figure 4.26. With a built-in pixel counter, users can also make sure that they get enough pixels across the subject to be able to identify it, whether it is a person's face or a license plate (for more information about determining resolution needs, see Chapter 17).

### 4.2.10 Lens quality

A light phenomenon called diffraction limits a lens' resolving power, which describes what the smallest size of details can be imaged. The bigger the lens, or the larger the aperture, the better



**Figure 4.25** Using the focus puller, the installer can make a rough focus adjustment before leaving the site.



**Figure 4.26** Through the camera’s user interface, the user can adjust the focus from any location. During the remote focusing process, the camera’s iris opens so that the back focus can be reset and then the focus fine-tuned. Simplified remote focus (a). Advanced remote focus (b).

the resolving power. All lens designs are the result of a number of compromises and are optimized for a particular task. The quality of the lens material, the coatings on a lens, and how the lens assembly is designed influence the resolution power a lens can provide.

Some lenses are made of glass and some of plastic. Although glass lenses are generally better, there is no guarantee that this is always the case. The properties and design of a lens are more important.

No lens is perfect, and all lenses create some form of aberration or image defects as a result of their limitations. The types of aberrations include the following:

- *Spherical aberration*: Light passing through the lens' edges is focused at a different distance than light striking near the lens' center.
- *Astigmatism*: Off-axis points are blurred in the radial or tangential direction. Focusing can reduce one at the expense of the other, but it cannot bring both into focus at the same time.
- *Distortion (pincushion and barrel)*: The image of a square object has sides that curve in or out.
- *Chromatic aberration*: The position of sharp focus varies with the wavelength.
- *Lateral color*: The magnification varies with wavelength.

#### 4.2.11 HDTV and megapixel lenses

High-resolution cameras with megapixel, HDTV, or Ultra-HD (4K, 8K) resolution put higher demands on lenses. Therefore, replacing a lens on a high-resolution camera needs some careful consideration. The main reason is that the pixels on megapixel sensors are much smaller than the pixels on Video Graphics Array (VGA) sensors and therefore demand a higher-quality lens.

It is not only the quality of the transparent material that matters but also the construction of the aperture (the iris). When it passes through the aperture, the light spreads out. This effect is called diffraction and cannot be avoided. It can be somewhat limited through careful construction of the aperture. Using a lens with P-Iris (see Section 4.2.5) optimizes the aperture in high-resolution cameras.

It is also critical to use the correct lens resolution. A 1/3-in. lens made for a VGA camera is often not suitable for a megapixel camera. This is because such a lens can resolve details with reasonable contrast only up to a certain number of lines per millimeter. It is best to match the lens resolution to the camera resolution to fully use the camera's capability and to avoid the effect of aliasing. The appropriate lens for a high-resolution camera with over 5 megapixels can be expensive. Read more about aliasing later in Section 4.5.8.

### 4.3 IMAGE SENSORS

When a network camera captures an image, light passes through the lens and falls on the image sensor. The image sensor consists of photosensitive diodes, called photosites, which register the amount of light that falls on them. The photosites convert the received amount of light into a corresponding number of electrons. The stronger the light, the more electrons are generated. The buildup of electrons in each photosite is converted into a voltage. An analog-to-digital (A/D) converter converts the voltage into digital data as picture elements. These picture elements are more commonly known as pixels. Once an image is formed, it is sent for processing in a stage that determines, among other things, the colors of each individual pixel that make up an image. Image processing is discussed in Section 4.5.

#### 4.3.1 Color filtering

Image sensors register the amount of light from bright to dark but no color information. Since these sensors are "color blind," a filter in front of the sensor allows the sensor to assign color tones to each pixel. Two common color registration methods are red, green, and blue (RGB) and cyan, magenta, yellow, and green (CMYG). RGB are the primary colors that, mixed in different combinations, can produce most of the colors visible to the human eye.

On their own, image sensors only register the amount of light, interpreted as shades of gray from white to black. Color can be registered using different methods. The most common RGB method takes advantage of the three primary colors, red (R), green (G), and blue (B), much like the human

eye does. By applying RGB color filters in a pattern over the pixels of an image sensor, each pixel can register the brightness of one of the three colors of light. For example, a pixel with a red filter over it records red light while blocking all other colors.

### COLOR AND HUMAN VISION

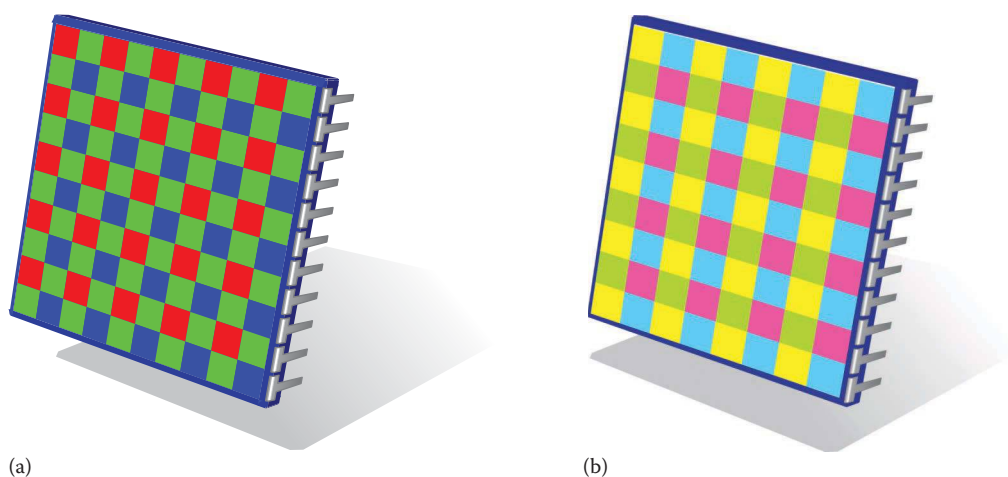
The human eye has three different color receptors: red, green, and blue. A camera is built to imitate the human visual system. This means that the camera translates light into colors that can be based not just on one light wavelength but a combination of wavelengths. When the wavelength of light changes, the color changes. Blue light has a shorter wavelength than red light. However, a combination of wavelengths or a single wavelength can be interpreted by the human visual system as the same color because the system is not able to tell the difference. Hence, yellow light, for example, could be either a single wavelength or a combination of several (e.g., green and red). Thus, there are certain different combinations of wavelengths that look the same.

The Bayer array, which has alternating rows of red–green and green–blue filters, is the most common RGB color filter (see Figure 4.27).

Since the human eye is more sensitive to green than to the other two colors, the Bayer array has twice as many green color filters. This also means that with the Bayer array, the human eye can detect more detail than if the filter uses the three colors in equal measures.

Another way to filter or register color is to use the complementary colors—cyan (C), magenta (M), and yellow (Y). Complementary color filters on sensors are often combined with green (G) filters to form a CMYG color array (see Figure 4.27). The CMYG system generally offers higher pixel signals due to its broader spectral band pass. However, the signals must then be converted to RGB since this is used in the final image, and the conversion implies more processing and added noise. The result is a reduction of the initial gain in signal-to-noise ratio, and the CMYG system is often not as good at presenting colors accurately.

The CMYG color array often is used in interlaced, CCD image sensors, whereas the RGB system primarily is used in progressive scan image sensors. The following section presents CCD and CMOS image sensor technologies. Section 4.4 gives more information about interlaced and progressive scan.



**Figure 4.27** Bayer array color filter placed over an image sensor (a). Cyan, magenta, yellow, and green color filter array (b).

### 4.3.2 CMOS and CCD technologies

Figure 4.28 shows CCD and CMOS image sensors. CCD sensors are produced using a technology developed specifically for the camera industry, whereas CMOS sensors are based on standard technology that is already extensively used in commonplace products such as PC memory chips. Modern CMOS sensors use a more specialized technology and the quality of the sensors has rapidly increased in recent years.

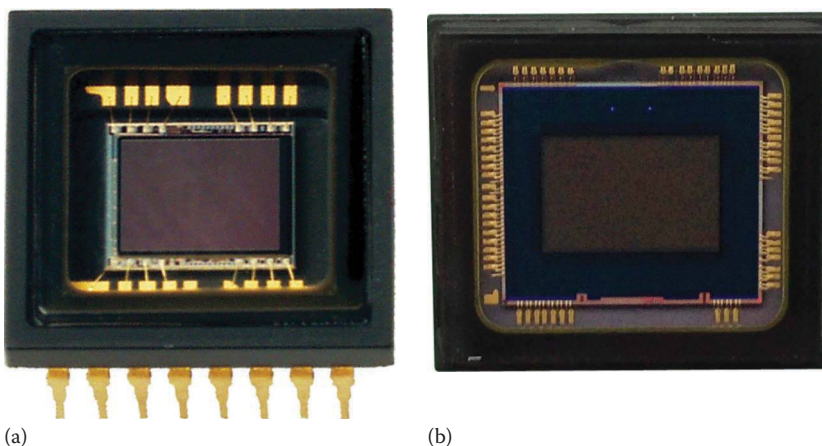
In a CCD sensor, every pixel's charge is transferred through a very limited number of output nodes (often one) in which the charges are converted to voltage levels, buffered, and sent from the chip as an analog signal. Then, the signal is amplified and converted into a set of numbers using an A/D converter outside the sensor. All the pixels in the sensor can be devoted to capturing light, and the uniformity of the output (a key factor in image quality) is high. The CCD sensor has no built-in amplifiers or A/D converters, so tasks are performed outside the CCD. CCD sensors are more expensive than CMOS sensors and more complex to incorporate into a camera. A CCD also can consume as much as 100 times more power than an equivalent CMOS sensor.

#### 4.3.2.1 CMOS technology

CMOS chips have several advantages. Unlike CCD sensors, CMOS chips contain all the logics needed to produce an image. They incorporate amplifiers and A/D converters, which lowers the cost of cameras. Every CMOS pixel contains conversion electronics. Compared to CCD sensors, CMOS sensors have better integration possibilities and more functions. However, this addition of circuitry inside the chip can lead to a risk of more structured noise, such as stripes and other patterns. CMOS sensors also have a faster readout, lower power consumption, higher noise immunity, and a smaller system size. Megapixel, HDTV, and 4K CMOS sensors are more widely available and less expensive than megapixel CCD sensors.

It is possible to read individual pixels from a CMOS sensor. This allows “windowing,” which means that parts of the sensor area can be read out, instead of the entire sensor area at once. This enables the delivery of a higher frame rate from a limited part of the sensor and that digital PTZ functions can be used. It is also possible to achieve multiview streaming, which lets one camera act as several virtual cameras. By allowing several cropped view areas to be streamed simultaneously from the sensor, one camera can simulate the views and tasks that would otherwise require multiple individual cameras.

While CMOS sensors were unusual in video surveillance cameras some 10 years ago, they have now come to dominate the market thanks to their improved performance, higher resolutions, and lower cost. Much of this development was driven by the demands of the consumer industry.



**Figure 4.28** A charge-coupled device sensor (a) and a complementary metal-oxide semiconductor sensor (b).



### 4.3.3 More about image sensors

Beyond size, resolution, and sensor type, there are several other characteristics that differentiate sensors:

- Pixel size
- Light sensitivity of a pixel (including fill factor)
- Maximum signal-to-noise ratio
- Dynamic range
- Fixed-pattern noise

Both the size of the image sensor and the size of the individual pixels affect image quality. Most network cameras usually use  $\frac{1}{4}$  or  $\frac{1}{3}$  in. image sensors with dimensions of no more than  $4.8 \times 3.6$  mm. A larger-sized image sensor can contain many more pixels than a smaller-sized sensor and can therefore deliver higher-resolution images and greater detail. When looking at image sensors used in megapixel sensors, always pay attention to the pixel count and pixel size. Two similar-sized image sensors that differ in pixel count and pixel size will likely produce different resolutions and have different levels of light sensitivity. See Section 4.3.4 for more information on megapixel sensors.

The bigger the pixels in an image sensor, the better they are at storing electrons generated from exposure to light. Generally, a larger pixel brings a larger maximum signal-to-noise ratio, effecting in images that are less noisy in highlights. Also, a larger pixel often means a higher fill factor and therefore a higher light-sensitivity level. The fill factor is the ratio of the area devoted to light gathering compared with the total area, which includes the area devoted to circuitry within a pixel. A sensor with pixels completely devoted to light gathering has 100% fill factor. Each pixel on a CMOS sensor has circuitry, and therefore the sensor's fill factor is less.

An image sensor also affects the dynamic range, that is, the range from the maximum useful level to the lowest noise level. To be able to capture both dark and bright objects in the same scene without showing too much noise, the camera's image sensor must have a high dynamic range (HDR). Many natural scenes have a rather high range of brightness levels, and sometimes it is difficult for a camera to handle them. Typical examples are indoor pictures with a bright window or an outdoor scene with a dark foreground and a bright sky. There are various techniques that enable a camera to go beyond the limited dynamic range of a typical sensor. They usually work in such a way that individual pixels are exposed or treated differently to reduce noise. Without such techniques, a higher gain on dark pixels also would amplify noise and give a lower-quality image.

The quality of an image sensor also is determined by how much fixed-pattern noise it has. Fixed-pattern noise is noise that has a pattern that does not change over time. It is caused by nonuniformity of the pixels on an image sensor and by electrons from heat generated by the sensor. It is mostly noticeable during long exposures.

### 4.3.4 HDTV and megapixel sensors

Megapixel and HDTV technology enable network cameras to deliver higher-resolution video images than analog closed-circuit television (CCTV) cameras. The higher resolution greatly improves the possibility to see details and to identify people and objects, which is key in video surveillance applications. Megapixel sensors are fundamental to HDTV, Ultra-HD, megapixel, and multi-megapixel cameras and can be used to provide extremely detailed images and multiview streaming.

Megapixel CMOS sensors are more widely available and generally less expensive than megapixel CCD sensors, even though there are plenty examples of very costly CMOS sensors.

It is difficult to make a fast megapixel CCD sensor, which of course is a disadvantage as it adds to the challenges of building a multi-megapixel camera based on CCD technology.



Assuming that the sensor technology and characteristics are equal, the sensor size should increase along with the increase in pixel quantity to have the same light sensitivity. Still, many sensors in multi-megapixel cameras (5 megapixels or more) are about the same size as 720p sensors. This means that the size of each pixel in a multi-megapixel sensor is smaller than the pixels in a 720p sensor. Because its pixels are smaller, a multi-megapixel sensor is typically less light sensitive per pixel than a 720p sensor. However, size is not everything. Quality also matters. Technology is rapidly evolving, so a modern small pixel may be better than a dated large pixel. Therefore, a smaller sensor may well be more light sensitive than a larger one. Again, testing and evaluating performance is the only way to know for sure that the sensor and other camera components are matched in such a way that they meet the expectations.

Megapixel sensors can be used innovatively in panoramic cameras to provide high-quality, full-overview images, or close-up images using instant zoom without moving camera parts. For more details on this subject, see Section 3.4.

## 4.4 IMAGE SCANNING TECHNIQUES

Interlaced scanning and progressive scanning are the two techniques that are currently available for reading and displaying information produced by image sensors. Interlaced scanning is mainly used in CCDs. Progressive scanning is used in both CCD and CMOS sensors. Network cameras can make use of either scanning technique, while most analog cameras are based on standards that can only use the interlaced scanning technique to transfer and display images.

### 4.4.1 Interlaced scanning

Interlaced scanning was originally introduced as a way to improve the image quality of a video signal without consuming additional bandwidth. The method soon became universal in traditional, analog television sets. To put it simply, the technique splits each frame into two fields. The scanning starts at the top-left corner and sweeps all the way to the bottom-right corner, skipping every alternate row on the way. This reduces the signal bandwidth by a factor of two, allowing for a higher refresh rate, less flicker, and better portrayal of motion.

There are a number of downsides to interlaced video, including the following:

- *Motion artifacts*: If objects are moving fast enough, they will be in different positions when each individual field is captured. This may cause motion artifacts.
- *Interline twitter*: A shimmering effect shows up when the subject being filmed contains very fine vertical details that approach the horizontal resolution of the video format. It is the reason news anchors would traditionally avoid using striped clothing.

All analog video formats and some modern HDTV formats are interlaced. The artifacts or distortions created through the interlacing technique are not very noticeable on an interlaced monitor. However, when interlaced video is shown on progressive scan monitors such as computer monitors, which scan lines of an image consecutively, the artifacts become noticeable. The artifacts, which can be seen as “tearing,” are caused by the slight delay between odd and even line refreshes as only half the lines keep up with a moving image while the other half wait to be refreshed. It is especially noticeable when the video is stopped and a freeze frame of the video is analyzed.

### 4.4.2 Deinterlacing techniques

To show interlaced video on computer screens and reduce unwanted tearing effects, different deinterlacing techniques can be used. The problem with deinterlacing is that two image fragments, captured at different times, must be combined into an image suitable for simultaneous viewing.

There are several ways to limit the effects of interlacing, including the following:

- Line doubling
- Blending

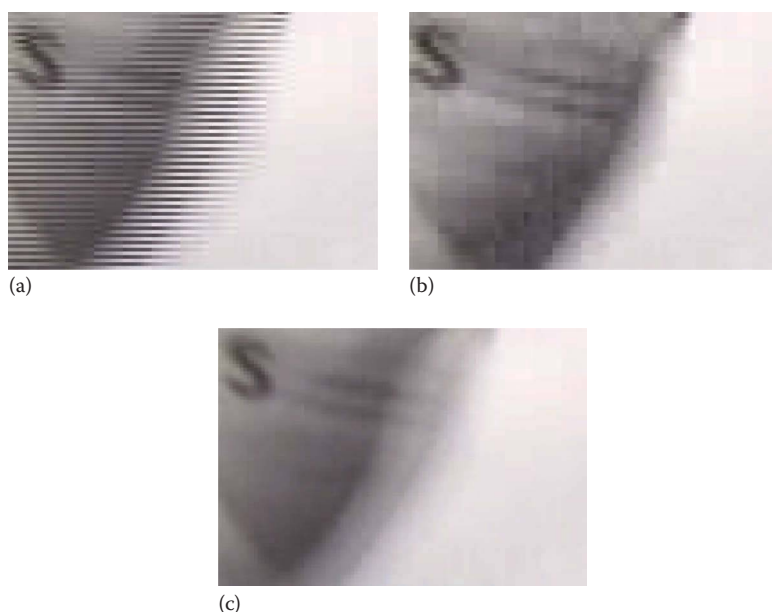
- bob deinterlacing
- Discarding
- Motion adaptive deinterlacing

Line doubling removes every other field (odd or even) and doubles the lines of each remaining field (consisting of only even or odd lines) by simple-line doubling or, even better, using interpolation. This results in the videos having effectively half the vertical resolution, scaled to the full size. Although this prevents accidental blending of pixels from different fields, called the comb effect, it causes noticeable reduction in picture quality and less smooth video (see Figure 4.29b,d).

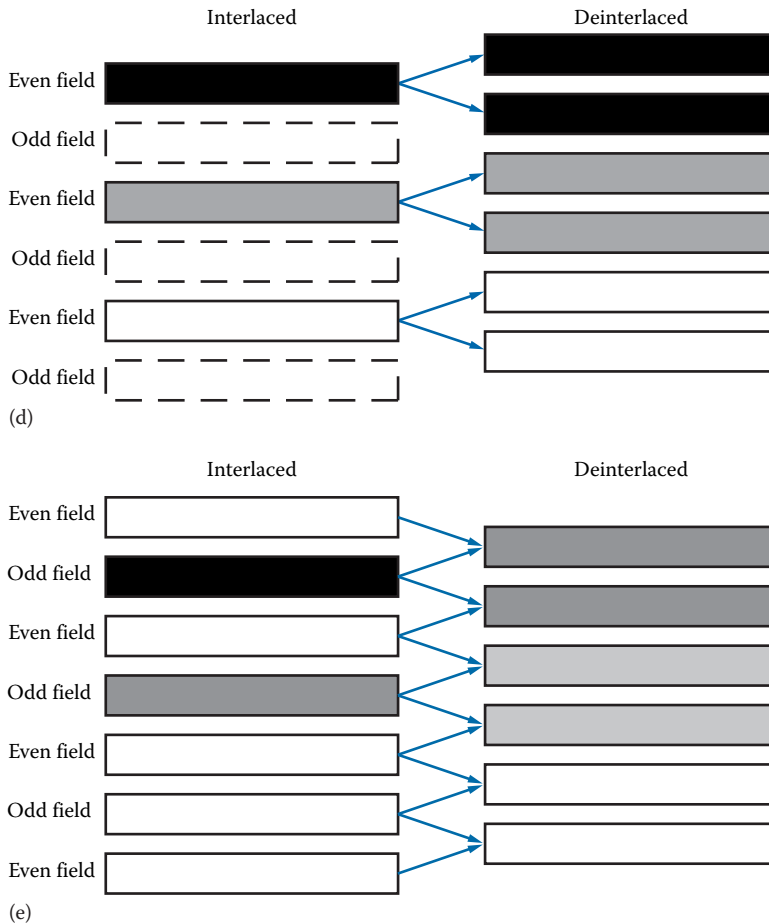
Blending mixes consecutive fields and displays two fields as one image. The advantage of this technique is that all fields are present, but the comb effect may be visible because the two fields, which are captured at slightly different time frames, are simply merged together. The blending operation may be done differently, giving more or less loss in the vertical resolution of moving objects. This often is combined with a vertical resize so that the output has no numerical loss in vertical resolution. The problem with this technique is that there is a loss in quality because the image has been downsized and then upsize. The loss in detail makes the image look softer, and the blending creates the ghosting artifacts (see Figure 4.29c,e).

In bob deinterlacing, the fields are “bopped” up and down, whereby each field is made into a full frame. Each odd frame is pushed down by half a line and each even frame is pushed up by half a line. This method requires one to know which field, whether odd or even, that should be displayed first. The total number of frames for the video is doubled, and to play the video at the right speed, the frame rate is doubled. This technique leads to some line flickering and requires more computer power to play and store.

Discarding is a trivial method that removes interlacing artifacts by scaling down the video horizontally and throwing out every other field and only keeping half of the information. The video size is smaller, and only half of the temporal resolution is kept, but no artifacts are visible.



**Figure 4.29** Original interlaced image shown on a progressive scan monitor (a). Image deinterlaced with line doubling (b). Image deinterlaced with blending (c). *(Continued)*



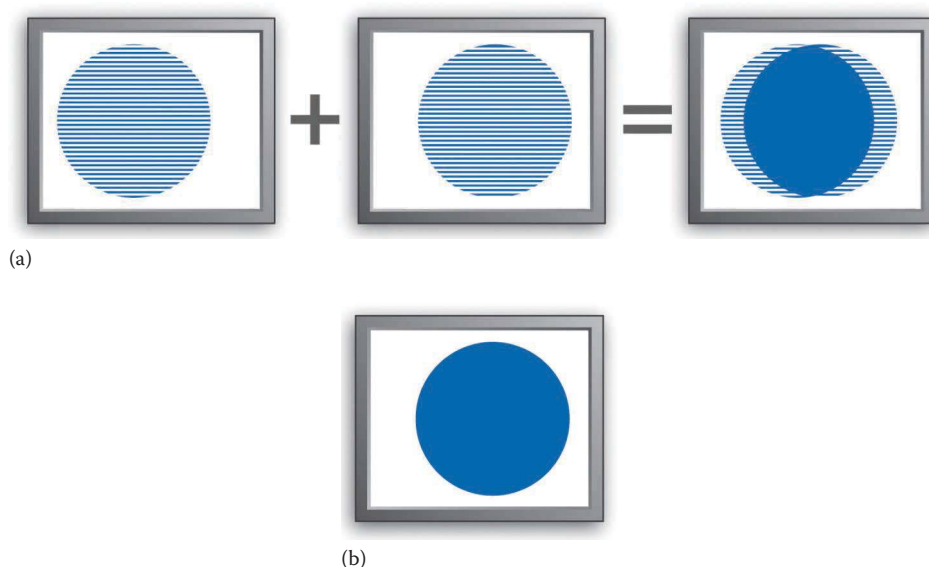
**Figure 4.29 (Continued)** Line doubling technique (d). Blending technique (e).

Motion adaptive deinterlacing is a more advanced technique. It incorporates the previously described technique of blending or averaging, together with a calculation for motion. If there is no motion, the two consecutive fields are combined to form a complete frame, and full resolution, as well as sharpness, is obtained. When motion is detected, vertical resolution is compromised as before. In pixel-based motion adaptive deinterlacing, the nonmoving parts of an image are shown in full resolution, while pixels that cause artifacts in the moving parts of the image are discarded. Lost data at the edges of moving objects then can be reconstructed using multidirection diagonal interpolation. This technique requires a great amount of processing power.

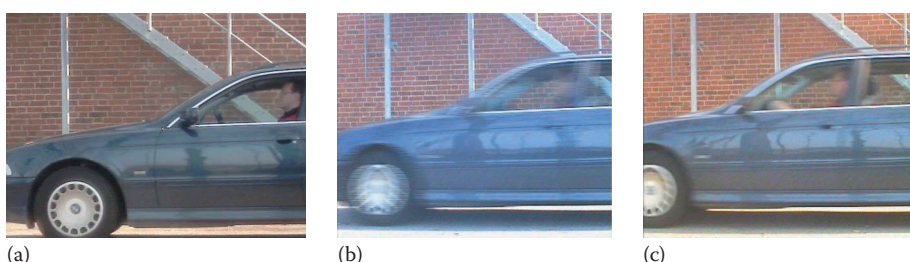
Interlaced scanning has served the analog camera, television, and VHS video world very well for many years and is still the most suitable technique for some applications. However, with the development and rapid adoption of TVs and displays using LCD, plasma, and LED technology, progressive scan has become dominant (see Figure 4.30).

#### 4.4.3 Progressive scanning

The progressive scanning technique captures, transmits, and displays all lines in the image in a single frame. Scanning is done line by line, from top to bottom. In other words, captured images are not split into separate fields like in interlaced scanning, so there is essentially no flickering effect. Therefore, moving objects are better presented on computer screens using the progressive scan technique. In a video surveillance context, progressive scanning can be critical to identify the details of a moving subject such as a person who is running away.



**Figure 4.30** On the left, an interlaced scan image shown on a progressive (computer) monitor. Odd lines. In the middle, even lines, 17/20 milliseconds (NTSC/PAL) later. On the right, freeze frame on moving dot using interlaced scan (a). Freeze frame on moving dot using progressive scan (b).



**Figure 4.31** Comparison between progressive, interlaced, and 2CIF-based scanning techniques. In these examples, the cameras are using the same lens and the car has the same speed (20 km/15 miles per hour): Progressive scan is used in network cameras, full size  $640 \times 480$  (a). Interlaced scan is mostly used in analog CCTV cameras, full size  $704 \times 576$  (b). 2CIF-based scan is used in digital video recorders, full size  $704 \times 240$  (PAL) (c).

When a camera captures a moving object, the sharpness of a still image depends on the technology used. Compare the JPEG images in Figure 4.31, which are captured by three different cameras using progressive scan, 4CIF interlaced scan, and 2CIF, respectively. For more information on resolution, see Section 4.6.

Note the following in the images in Figure 4.31:

- All image systems produce a clear image of the background.
- With interlaced scan images, jagged edges result from motion.
- In the 2CIF sample, motion blur is caused by lack of resolution.
- Only progressive scan makes it possible to identify the driver.

With progressive scanning, single frames from a video sequence can be used to make paper copies with almost photographic quality. This is a strong benefit. It can even be crucial if the material is to be used as evidence in a court of law.

## 4.5 IMAGE PROCESSING

The lens and sensor are key camera components and determine much of the quality of an image. The quality can be further improved by the image processor, which is built into the network camera. The image processor can use a number of techniques and adjust or apply different parameters to process the image, including the following:

- Control of exposure time, aperture, and gain
- Backlight compensation and WDR
- Bayer demosaicing that converts raw data into a color image
- Noise reduction
- Color processing such as saturation and white balance
- Image enhancement such as sharpening and contrast

### 4.5.1 Exposure

Thanks to the human eye's ability to adapt to different lighting conditions, we can see and navigate in very dark environments and extremely bright environments. A camera also must be able to cope with changes in brightness. To do this, the camera must find the correct exposure.

In a camera, there are three basic controls that are used to achieve the appropriate exposure and therefore the ideal image quality:

1. *Exposure time*: How long the sensor is exposed to incoming light
2. *Aperture or f-stop*: How big the iris opening is that lets light through to the sensor
3. *Gain*: Image level amplifier that can make the images look brighter

Exposure time, aperture, and gain are often set automatically by the camera, but many cameras also have direct or indirect tools for manual adjustment. By adjusting one or more of these settings, an image of a scene will appear relatively unchanged even when the lighting at the scene changes.

The two images in Figure 4.32 are taken under different illumination conditions. Thanks to auto-exposure, the images appear relatively similar. The images are taken in the same room, but in Figure 4.32a the figure is standing on top of a cabinet and in Figure 4.32b the figure is standing inside the same cabinet, so the lighting is different.



(a)



(b)

**Figure 4.32** Even though they have different illuminance, through autoexposure two images of the same scene can look the same. In (a), the figure is illuminated by afternoon sunlight coming from the right through a window about 50 cm (20 in.) away. In (b), the figure is only illuminated by indirect light.

Be aware that increasing the gain of the image not only increases the image luminance but also increases the noise. Adjusting the exposure time or using a smaller f-stop (larger iris opening) is therefore preferred. However, sometimes the iris is fully open and the exposure time required is longer than the time between two frames. For example, if the frame rate is 30 fps and the exposure time is 1/15 second, the exposure time is twice as long as the time between each frame. Then a decision has to be made: lower the frame rate or increase the gain? Which to sacrifice depends on the application requirements and on what the priorities are.

### 4.5.2 Backlight compensation

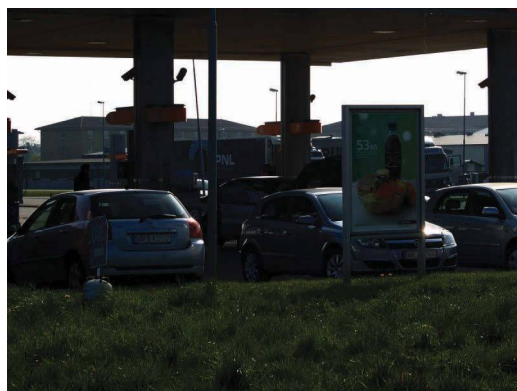
Although a camera's automatic exposure controller tries to get the brightness of the image to appear as the human eye would see the scene, it can be easily fooled. Strong backlight is difficult for a camera to handle (Figure 4.33).

In the case with strong backlight, such as in Figure 4.33a, the camera believes that the scene is very bright. So the camera makes the iris opening smaller or reduces the exposure time, which results in a dark image. A feature called backlight compensation helps prevent this from happening. It strives to ignore limited areas of high illumination as if they were not present. The resulting image (Figure 4.33b) makes it possible to identify all the objects in the scene. The bright areas, however, are overexposed. Without backlight compensation, the image would be too dark, and color and details would be lost in the foreground. Such a situation also might be dealt with by increasing the dynamic range of the camera, as discussed in the next section.

In addition to dealing with limited areas of high illumination, a network camera's automatic exposure controller must decide what area of an image to base the exposure value on. For example, the foreground (such as a bus, a city square, or railway tracks) can hold more essential information than the background that often takes up a part of the upper half of an image (such as the sky or a line of trees). The less important areas of a scene should not determine the overall exposure. The camera designer's solution should be to divide the image into subimages and assign different weights for the exposure algorithm (Figure 4.34). In advanced network cameras, the user is able to select which of the predefined areas should be more correctly exposed. The position of the window or exposure area can be set to center, left, right, top, or bottom.

### 4.5.3 WDR

Assigning different weights for different sections of an image is one way to determine exposure values. But there are many surveillance situations that present lighting challenges that cannot be solved by changing the exposure time or creating exposure zones.



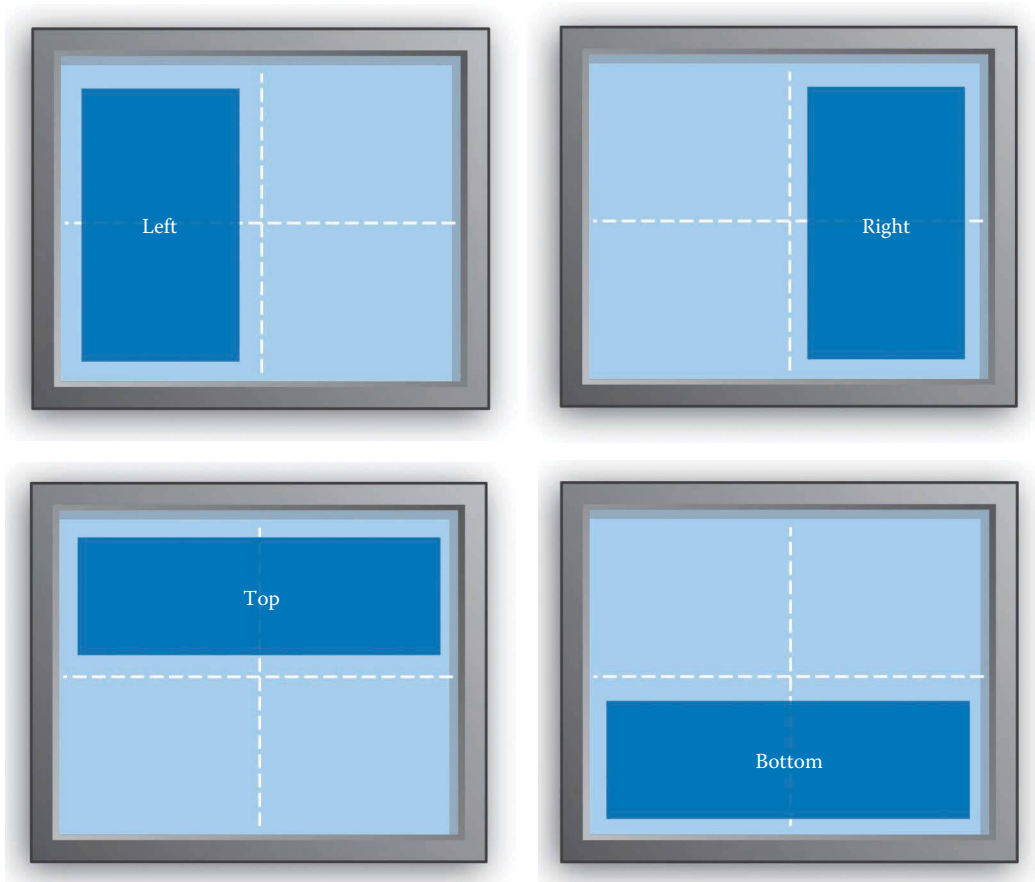
(a)



(b)

**Figure 4.33** Strong backlight, without backlight compensation (a). With backlight compensation applied, limited areas of high illumination can be ignored (b).





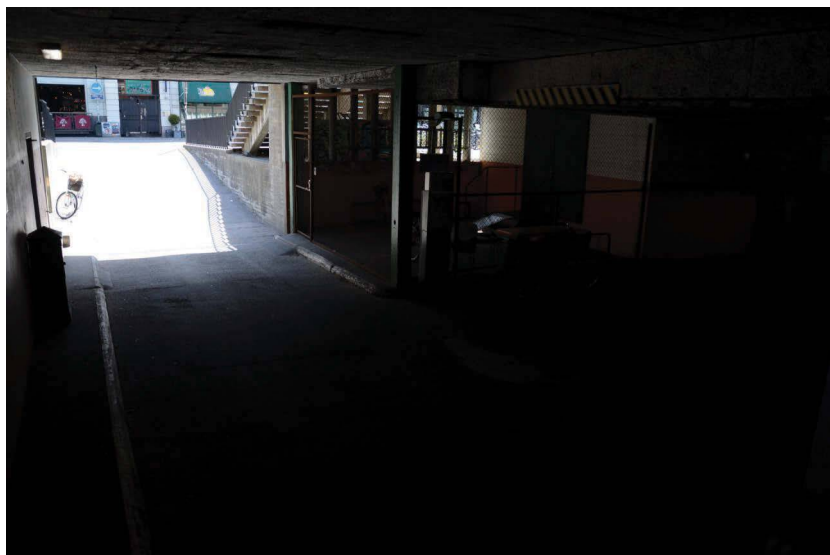
**Figure 4.34** Assigning different weights to sections of an image to better determine which area should be more correctly exposed.

An overcast day with few shadows has a low dynamic range. That is, there are no deep blacks and no extreme highlights. But on sunny days where there are really distinct shadows, there could be a big difference between the brightest and darkest areas. This is called a WDR, also known as HDR. In nature, there are dynamic ranges that extend further than the camera or the human eye can perceive. In these situations, the camera has to make a decision about which area to expose properly.

Typical situations include the following:

- Entrance doors with daylight outside and a darker indoor environment. This is very common in retail, banking, and office environments.
- Vehicles entering a parking garage or tunnel, also with daylight outside and low light levels indoors.
- In city surveillance, transportation, perimeter surveillance, and other outdoor applications where parts of the scene are in direct sunlight and other parts are in deep shadows.
- Vehicles with strong headlights, driving directly toward the camera.
- Environments with lots of reflected light such as office buildings with many windows and shopping malls.

WDR imaging is a method for restructuring images with full scene content so that people, objects, vehicles, and events can be reliably identified in situations where there is a wide range of lighting conditions. Without WDR imaging, the camera would produce an image where objects in the dark area of the scene would barely be visible (Figure 4.35).



(a)



(b)

**Figure 4.35** A parking garage where the camera is placed inside the garage. One image is underexposed (a), and the other image is overexposed (b).

Clearly both images lack information from the full scene. A good WDR surveillance camera can capture both these two extremes in one image, that is, clearly show details in the well-lit entrance as well as the dark shadows inside the parking garage (Figure 4.36).

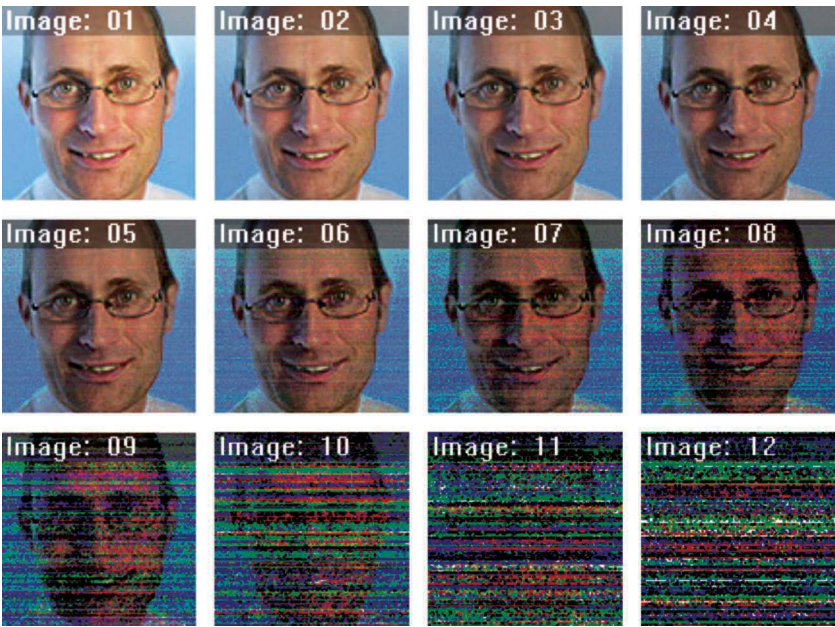
#### 4.5.3.1 Measuring dynamic range

Not all situations need a WDR camera. Using a low dynamic range camera that is configured to avoid clipping the highlights, the following image examples show how the noise increases when a person is moved into darker and darker shade. Together, Figure 4.37 and Table 4.6 give an indication of when WDR technology must be used.

The lower the image quality and contrast, the less useful the image is in, for example, forensic investigations. Look at the images in Figure 4.37 and ask yourself, which image is good enough to fulfill its



**Figure 4.36** The same parking garage, but this image has been captured using wide dynamic range (WDR) technology (capture WDR). The areas that were underexposed are brighter and the areas that were overexposed have been leveled.



**Figure 4.37** Images of subject taken by a standard camera with default settings. As the light darkens, the noise increases.

**Table 4.6** Type of scene and its typical illumination ratio

Type of scene	Example	Illumination ratio
(a) Sunlight/shadow	A typical example is a sunlit train station where a part of the platform shielded from the sun.	1:20
(b) Window illumination indoor	Lobby with large windows.	1:200
(c) Dark indoor scene with opening	Typical warehouse scene with door or opening to the sunlight outside.	1:2000

purpose? In scenes where there is a good balance between shade and lighting, there is a better chance the images will be of high enough quality to meet the requirements. If the scene is closest to type (a), sunlight/shadow, you could expect a standard camera to deliver images with a quality equal to image 06. Image 06 is good enough for identification and many forensic purposes. As the lighting conditions worsen, you must lower your expectations on image quality accordingly and tolerate a higher noise level. If the scene is closest to type (b), window illumination, a standard camera might deliver an image quality equal to image 09. Note that image 09 provides little more than detection level. In conditions like type (c), a dark indoor scene, you can expect noise levels equal to or higher than image 12. Remember, this is only an indication since standard cameras obviously vary in quality.

As dynamic range varies greatly, the dynamic range capability of a camera is often presented in the logarithmic unit of decibels (dB). It expresses the ratio of radiance between the brightest and the dimmest object that can be captured by the camera. Note that this is not the same as the illumination ratio used in Table 4.6. If the radiance ratio is 1000:1, the dB value is 60 dB. The value is calculated as the logarithm (in this case 3) of the ratio times 20, just like one calculates the amplitude ratio of voltage:

$$Ratio_{dB} = 10 \times \log_{10} \left( \frac{V_1^2}{V_0^2} \right) = 20 \times \log_{10} \left( \frac{V_1}{V_0} \right)$$

$$Ratio_{dB} = 20 \times \log_{10} \left( \frac{1000}{1} \right) = 20 \times 3 = 60 \text{ dB}$$

The dimmest detectable level can be defined as the noise floor of the sensor pixel. Any signal below this level is drowned in noise. With this definition, a good image sensor can reach a dynamic range of about 70 dB. Through techniques such as multiframe exposures, some sensors can increase the upper detection limit. This extends the dynamic range to more than 100 dB, but this is not necessarily always the best WDR solution.

Although some modern surveillance cameras use sensors with extended dynamic range that allows them to better handle difficult scenes, the dB number cannot fully describe the WDR capacity. The decibel unit is just an approximation of the dynamical capabilities of the camera. As is the case with lux ratings, manufacturers can choose to use their own measuring methods, which makes datasheet comparisons unreliable. To know the capabilities of the cameras fully, it is best to test them on-site.

#### 4.5.3.2 Types of WDR

There are several types of WDR. Each type is designed to address different types of situations and use cases. The types of WDR include:

- Contrast WDR
- Capture WDR
- Forensic WDR

Contrast WDR (also known as WDR—dynamic contrast) allows the camera sensor capture an image with a higher bit depth than what the camera ultimately can send out. The bit depth is an internal property that translates to dynamic range. Contrast WDR performs an advanced tone mapping, where some brightness levels are dropped to decrease the bit depth to a format that a computer screen can handle. Tone mapping looks at both the darkest and the brightest parts. The result is an image that has more details in both ends.

There are two types of tone mapping. With global tone mapping, all pixels are handled in the same way, which means that the same levels are removed everywhere in the image. With local tone



mapping, individual decisions are made for different regions in the image to decide which levels to remove. Local tone mapping requires much more processor power but gives a better result.

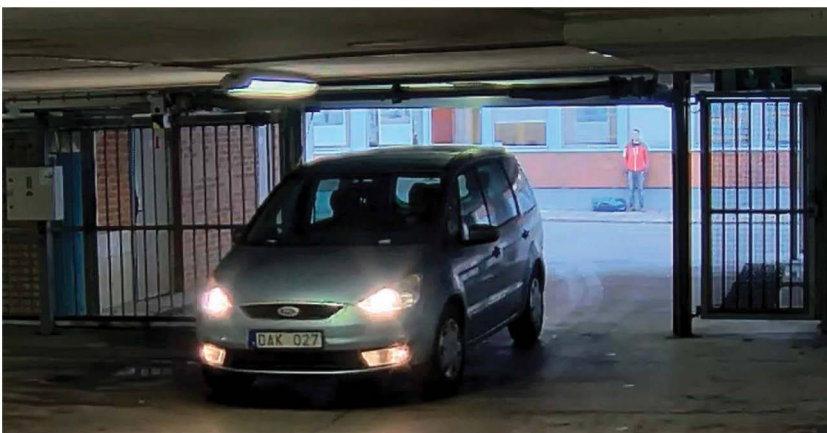
Capture WDR (also known as WDR—dynamic capture) takes several images with different exposure levels for each image. These pictures are then combined in a composite, where both the brightest and the darkest parts are kept, resulting in an image with more clarity and sharpness. However, this image has a higher bit depth than a computer screen can handle. Therefore, tone mapping needs to be applied just as with contrast WDR. Capturing several images in the time span normally used for one image requires an extremely fast and sensitive sensor. Still, the output is much better than with contrast WDR (Figure 4.38).

Forensic WDR (also known as WDR—forensic capture) applies advanced algorithms that optimize image quality especially tuned to forensic purposes. The algorithms lower noise levels and increase the image signal to display every detail in the best possible way. The method also includes the ability to seamlessly transition between WDR and low-light mode. The resulting images often look different from what we have learnt to expect from television broadcasts and still image photography, but this WDR technology makes it easier to detect and identify critical details in a scene (Figure 4.39).

Image results from a WDR camera will differ depending on such aspects as the complexity of the scene and amount of movement. As with any video surveillance situation, the most important questions are as follows: What do you want to see? And how do you want to present the captured image?

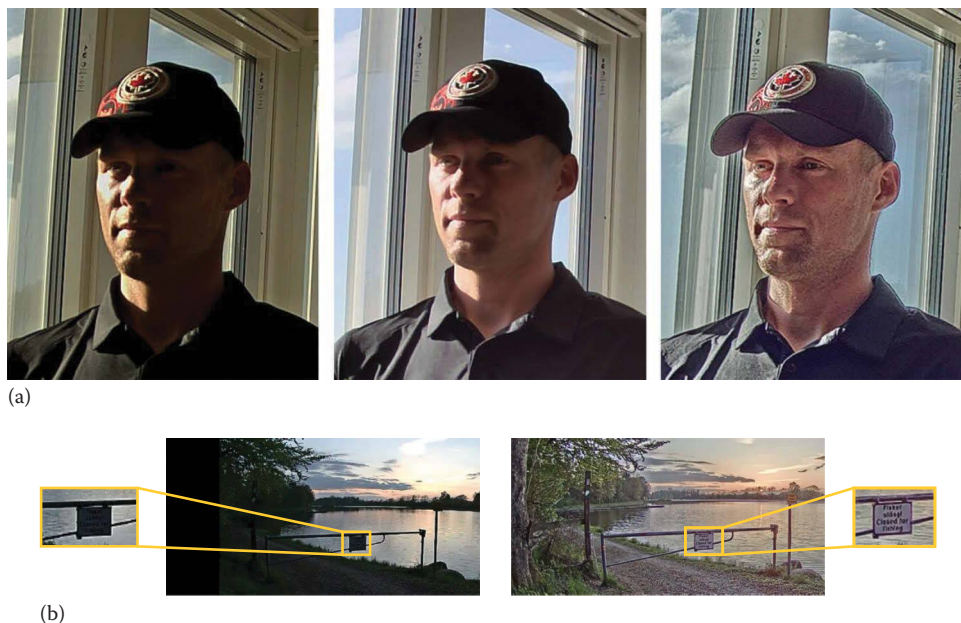


(a)



(b)

**Figure 4.38** A parking garage where the camera is placed inside the garage. One image is taken without wide dynamic range (WDR) imaging (a), and the other is taken with capture WDR (b).



**Figure 4.39** Two scenes that illustrate the effects of forensic wide dynamic range (WDR). With forensic WDR, shadows and highlights are leveled, creating a high-contrast image suitable for forensic investigations: With capture WDR (a –left and middle), the contrast in the bright and dark parts are enhanced, making it possible to see details that otherwise would have been too dark. However, forensic WDR (a –right) is even better at giving the type of differentiating details, such as chin dimples, birthmarks, and tattoos, that make a positive identification possible. Without WDR (b –left), the light from the sunset creates a strong backlight that wipes out all details, making the sign completely black. With forensic WDR (b –right), it is suddenly possible to make out the letters on the sign (the first two words are in Swedish).

HDR technologies typically use different exposures for different objects within a scene and employ ways to display the results. However, the screen used to display the video may also have limited dynamic range.

Although HDR technologies solve some problems, it sometimes introduces others. Here are some examples:

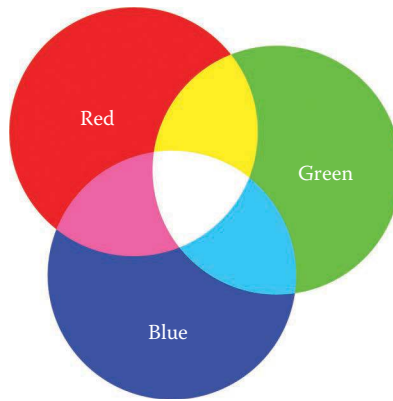
- Noise can be very different in different regions of an image. In particular, dark regions may contain very visible levels of noise.
- Pixels between two different exposure regions may show visible artifacts. This can be seen in images with high dynamic content and many different levels of lighting at the same time.
- Different exposure regions may have been allocated a dynamic range that is too low, making every part of the image bad.
- Colors may look bleak.

#### 4.5.4 Bayer demosaicing

After the sensor captures the raw image, the camera processes it to produce a high-quality color image.

In a process called demosaicing, an algorithm is used to translate the raw data from the image sensor into full-color information at each pixel. Because each pixel only records the illumination behind one of the color filters, it needs the values from the other filters—interpolated from neighboring pixels—to calculate the actual color of a pixel. For example, if a pixel with a red filter registers a bright value and the neighboring green and blue pixels are equally bright, then the camera’s processor determines that the color of the pixel with the red filter actually must be white (Figure 4.40).





**Figure 4.40** When red, green, and blue light are mixed in equal amounts, white is the result.

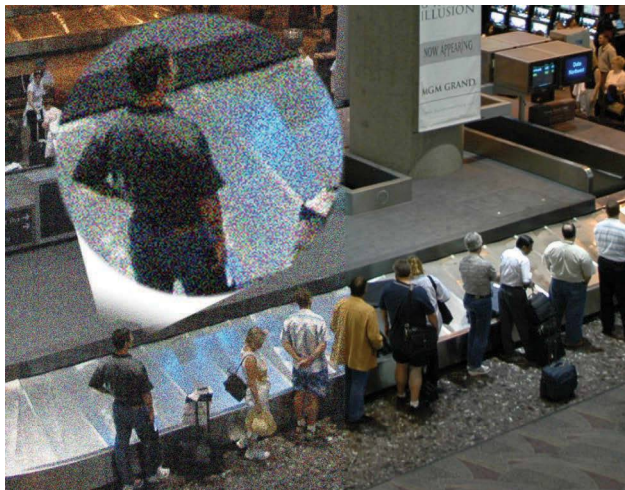
### 4.5.5 Noise

Because no sensor is perfect, all cameras have an uncertainty with regard to the pixel values produced. This is known as random noise. It means that even if the scene and the illumination are the same, the same pixel will not give the exact same value each time it is read out. Sometimes, random noise appears as banding when entire rows are affected. Every sensor also has a few bad, nonworking pixels.

There are also differences between individual pixels within a typical sensor. This means that adjacent pixels exposed to exactly the same light generally do not respond exactly the same way. This appears as a fixed-pattern noise that does not change over time. Some fixed-pattern noise changes with temperature and exposure time and is therefore more pronounced in hot environments or when the exposure time increases at night.

Reducing noise is a key task in a video surveillance camera. Part of the noise is generated within the camera, but part of it actually is due to the nature of light itself and therefore affects all cameras. The latter is mostly visible in bright daylight images of, for example, blue skies. Noise is lower in cameras that use sensors with larger pixel sizes, which can collect more light in each pixel.

An image taken in low light might appear grainy or have specks of color because the noise is amplified together with the signal (Figure 4.41). This is true for both random noise and fixed-pattern noise. In a video camera, fixed-pattern noise can be recognized by the constant position of the specks, but it cannot be distinguished from random noise in a still image camera.



**Figure 4.41** Half of the image is shown with noise. Noise appears as specks of colors that distort the image.

Noise can be reduced by various filtering techniques that replace flawed pixels with new values calculated from neighboring pixels. Most cameras include one or several filters in their processing. Most filters, however, have visible side effects that appear as reduced resolution, motion blur, or other artifacts.

Modern network cameras with chips that offer high processing capacity are well equipped to analyze and reduce noise levels, even in high resolution and full-frame-rate conditions. One technique for minimizing noise is built on spatial processing and another on temporal processing.

- *Spatial processing* analyzes a single image frame to find pixels that are very different in color or intensity from their surrounding pixels.
- *Temporal processing* compares consecutive image frames to find artifacts in the images that are not static over time and can be regarded as potential noise.

### 4.5.6 White balance

Once color interpolation is done for an image, the image processor adjusts the white balance in the image to make sure that the image has the right color balance (Figure 4.42).

To achieve the right color, neutral (black, gray, white) colors in a scene should stay neutral in the resulting image, regardless of the illumination. In a network camera, white balance can be set by choosing between an automode and a selection of presets tailored to indoor and outdoor environments. With automatic white balance, the camera uses two or three different gain factors to amplify the RGB signals.

### 4.5.7 Sharpening and contrast

Two tools that are used for fine-tuning of digital images are digital sharpening and contrast enhancement (Figure 4.43). Digital sharpening changes the edge contrast, not the resolution.

- *Sharpening* increases the local contrast by lightening the light pixels and darkening the dark pixels at the edges. But be aware that sharpening can also amplify noise.
- *Contrast enhancement* affects not only the edges but all pixels in the image equally. Contrast enhancement changes how the original pixel values are mapped or converted to values used by a display monitor.



(a)

**Figure 4.42** An image with a reddish tint (a).

(Continued)



(b)



(c)

**Figure 4.42 (Continued)** An image with a greenish tint (b). An image with white balance applied (c).

Fog, haze, smoke, and similar weather conditions have a special effect on images that can be corrected through contrast enhancement. Whole image areas lose their visibility and objects lose their contrast. Defogging automatically detects the foggy elements and tweaks them digitally to deliver a clearer image (see Figure 4.44). It is an advanced contrast enhancement algorithm that analyzes the image pixel by pixel. So the amount of tweaking varies throughout the scene and only affects the dim areas but not the areas that are already bright and clear. In some cases, defogging can make the difference to whether a vehicle is identified or not.

### 4.5.8 Aliasing

Sometimes a subject contains finer details than the size of the pixels in an image sensor. The sensor cannot detect such details because the pixel resolution is too low. To see fine details, the image scale must be increased with a telephoto lens, or a sensor with a higher resolution must be used.



**Figure 4.43** Sharpening versus contrast: An image before sharpening and changing of contrast levels (a). Contrast reduced – the darker areas become more visible and the lighter areas become less bright (b). Sharpening applied (c).

Repeated unresolved patterns, such as the herringbone pattern on tweed jackets, can cause problems for some cameras due to an effect called aliasing. The effect is due to poor resolution and appears as distortions in the form of unwanted, larger patterns that can even be colored in the image. The two images in Figure 4.45 are representations of the same pattern on a jacket as seen with two different image sensor resolutions. The image to the right has a pronounced aliasing effect.

Cameras translate continuous grades of tone and colors to points on a regular sampling grid. When details are finer than the sampling frequency, a good camera averages the details to avoid aliasing. This discards irresolvable details, but if too many details are filtered out, the image could look soft.

## 4.6 RESOLUTION

Resolution in an analog or digital world is similar, but there are some significant differences in how it is defined. Because analog video has its origin in the television industry, its images consist of lines (or television lines). In a digital system, an image consists of square pixels, which is another word for picture elements.



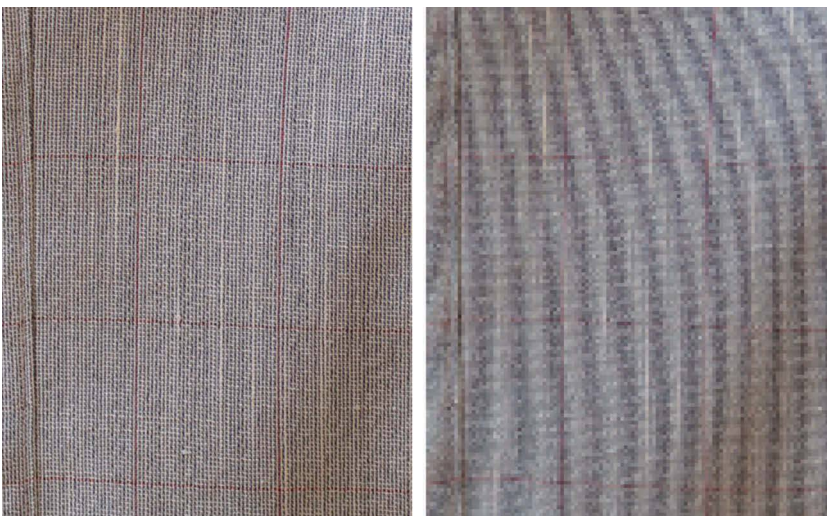


(a)



(b)

**Figure 4.44** Two images of a parking lot scene in foggy weather: defogging off (a) and defogging on (b).



**Figure 4.45** The same pattern on a jacket as seen with two different image sensor resolutions. Note pronounced aliasing effect in the image to the left.

4.6.1 NTSC and PAL resolutions

In North America and Japan, the NTSC standard is the predominant analog video standard, whereas in Europe the PAL standard is used. Both standards were established by the television industry.

NTSC has a resolution of 480 lines and uses a refresh rate of 60 interlaced fields per second (or 30 full frames per second). A new naming convention that defines the number of lines, scanning type, and refresh rate for this standard is 480i60, where “i” stands for interlaced scanning.

PAL has a resolution of 576 lines and uses a refresh rate of 50 interlaced fields per second (or 25 full frames per second). The new naming convention for this standard is 576i50. The total amount of information per second is the same in both standards.

Common Intermediate Format (CIF) is a standardized system for converting the NTSC and PAL standards to digital format. When analog video is digitized, the maximum number of pixels that can be created is based on the number of TV lines available for digitization. The maximum size of the digitized image is 720 × 480 pixels (NTSC) or 720 × 576 pixels (PAL). This resolution is also known as D1.

4CIF has 704 × 480 pixels (NTSC) or 704 × 576 pixels (PAL). 2CIF has 704 × 240 pixels (NTSC) or 704 × 288 pixels (PAL), which means that the number of horizontal lines is divided by two. In most cases, each horizontal line is shown twice when shown on a monitor. It is an effort to maintain correct ratios in the image and is called line doubling. It is also a way to cope with motion artifacts caused by interlaced scan. Section 4.4.2 discusses deinterlacing, and Figure 4.46 shows a comparison of the different NTSC and PAL image resolutions.

Sometimes, a quarter of a CIF image is used. This resolution is called Quarter CIF or QCIF for short.

Since the introduction of network cameras, it has been possible to design surveillance systems that are fully digital. This renders the limitations of NTSC and PAL irrelevant. Several new resolutions stemming from the computer and digital television industry have been introduced. They are world-wide standards and give better flexibility.

4.6.2 VGA resolutions

VGA is a graphics display system for PCs originally developed by IBM. The resolution is defined as 640 × 480 pixels, which is a format commonly used by nonmegapixel network cameras. VGA resolution

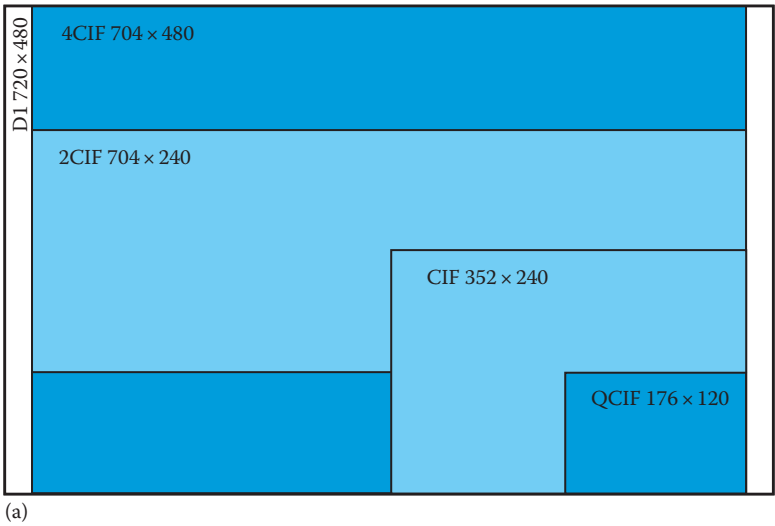
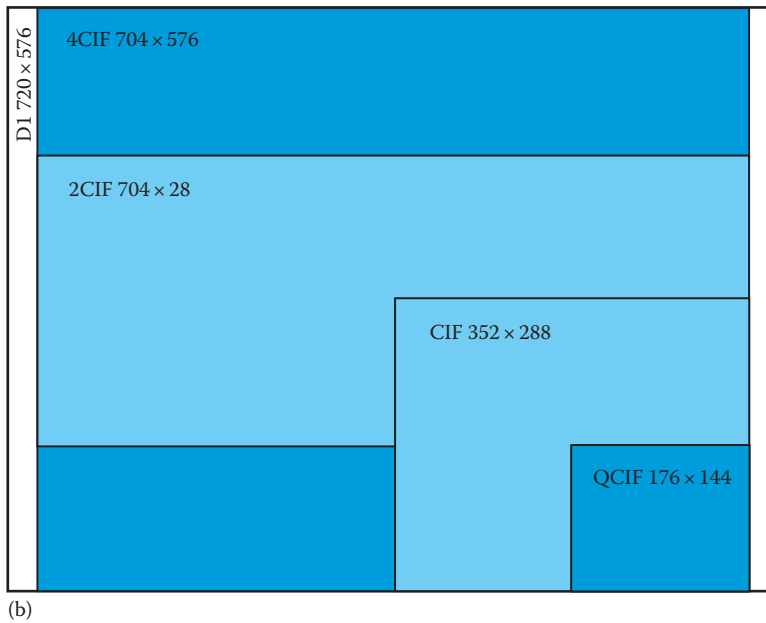


Figure 4.46 Different NTSC image resolutions (a).

(Continued)





**Figure 4.46 (Continued)** Different PAL image resolutions (b).

**Table 4.7** VGA resolution

Display format	Pixels
QVGA (SIF)	320 × 240
VGA	640 × 480
SVGA	800 × 600
XVGA	1024 × 768
4 × VGA	1280 × 960

is normally better suited for network cameras than CIF resolutions because the video will be shown, in most cases, on computer monitors with VGA resolutions or higher and not traditional TV monitors.

Quarter VGA (QVGA), with a resolution of 320 × 240 pixels, is very similar in size to CIF. QVGA is sometimes called Standard Interchange Format (SIF) and can be easily confused with CIF. Other VGA-based resolutions are SVGA (800 × 600 pixels), XVGA (1024 × 768 pixels), and 1280 × 960 pixels, which is four times VGA and provides megapixel resolution. For more about megapixel resolutions, see Section 4.6.4.

Table 4.7 summarizes the VGA resolutions.

### 4.6.3 MPEG resolutions

In the early days of MPEG-1 and MPEG-2, the number of resolutions was limited. They were also somewhat different and referred to as D1 or parts of D1. The actual MPEG compression methods did not impose the limitations, just the implementation limits and *de facto* standards. To read more about the history of MPEG, see section 6.2.2. Today, MPEG resolution (Figure 4.47) usually means one of the following resolutions:

- 704 × 576 pixels (PAL 4CIF)
- 704 × 480 pixels (NTSC 4CIF)
- 720 × 576 pixels (PAL or D1)
- 720 × 480 pixels (NTSC or D1)



**Figure 4.47** Different MPEG image resolutions.

**4.6.4 Megapixel resolutions**

A network camera that offers megapixel resolution uses a megapixel sensor to deliver an image that contains millions of pixels (one million or more). The more pixels a sensor has, the greater the potential it has for capturing finer details and for producing a higher-quality image. Megapixel network cameras are used in surveillance scenarios where details are critical, such as when people and objects need to be identified, when a larger area needs to be monitored, or when the scene needs to be divided into multiple view areas.

Table 4.8 and Figure 4.48 summarize some megapixel formats.

In the video surveillance industry, best practices have emerged regarding the number of pixels required for certain situations. For an overview image, the general opinion is that about 70–100 pixels are enough to represent 1 m (20–30 pixels per foot) of a scene. For situations that require detailed images, such as face identification, the demands can rise to as many as 500 pixels per meter (150 pixels per foot). For example, this means that if it is necessary to identify people passing through an area that is 2 m wide and 2 m high (6 ft<sup>2</sup>, 6¾ in.<sup>2</sup>), the camera must be able to provide a resolution of at least 1 megapixel (1000 × 1000 pixels).

Megapixel resolution is one area in which network cameras excel over analog cameras (Figure 4.49). The maximum resolution a conventional analog camera can deliver after digitizing

**Table 4.8** Megapixel formats

Display format	No. of megapixels	Pixels
SXGA	1.3	1280 × 1024
SXGA+ (EXGA)	1.4	1400 × 1050
UXGA	1.9	1600 × 1200
WUXGA	2.3	1920 × 1200
QXGA	3.1	2048 × 1536
WQXGA	4.1	2560 × 1600
QSXGA	5.2	2560 × 2048



**Figure 4.48** Different megapixel image resolutions.



**Figure 4.49** Megapixel camera view versus analog camera view.

the video signal in a digital video recorder or a video encoder is D1, which is  $720 \times 480$  pixels (NTSC) or  $720 \times 576$  pixels (PAL).

The D1 resolution corresponds to a maximum of 414,720 pixels, or 0.4 megapixels. By comparison, a common megapixel format is  $1280 \times 1024$  pixels, giving a resolution of 1.3 megapixels. This is more than three times higher than the resolution that analog CCTV cameras can give.

### 4.6.5 HDTV resolutions

HDTV provides up to five times higher resolution than standard analog TV. HDTV also has better color fidelity and a 16:9 format. Defined by Society of Motion Picture & Television Engineers (SMPTE), the two most important HDTV standards are SMPTE 296M that defines the HDTV 720p format and SMPTE 274M that defines the HDTV 1080 format.

#### *HDTV 720p*

- 1280 × 720 pixel resolution
- 0.9 megapixels
- High color fidelity (color spaces are defined by ITU-R BT.709)
- 16:9 aspect ratio
- Progressive scan
- Main compression standard being H.264, although H.265 can be used
- 25/30 and 50/60 Hz refresh rate, which corresponds to 25/30 and 50/60 fps (the frequency is country dependent)

#### *HDTV 1080p*

- 1920 × 1080 pixel resolution
- 2.1 megapixels
- High color fidelity (color spaces are defined by ITU-R BT.709)
- 16:9 aspect ratio
- Interlaced or progressive scan
- Main compression standard being H.264, although H.265 can be used
- 25/30 and 50/60 Hz refresh rate, which corresponds to 25/30 and 50/60 fps (the frequency is country dependent)
- Also known as Full HD

A camera that complies with the SMPTE standards indicates adherence to HDTV quality and should provide all the benefits of HDTV in resolution, color fidelity, and frame rate.

The HDTV standard is based on square pixels—similar to computer screens—so HDTV video from network video products can be shown on either HDTV screens or standard computer monitors. With progressive scan HDTV video, no conversion or deinterlacing technique needs to be applied when the video is to be processed by a computer or displayed on a computer screen. For more about deinterlacing, see Section 4.4.2.

Tables 4.9 and 4.10 list the basic HDTV image sizes in the European Broadcasting Union and NTSC countries, respectively.

**Table 4.9** Basic HDTV image sizes in the European Broadcasting Union

Size	Aspect ratio	Scan	Frame rate (fps/Hz)	Label
1280 × 720	16:9	Progressive	50	720p50
1920 × 1080	16:9	Interlaced	25 <sup>a</sup>	1080i50
1920 × 1080	16:9	Progressive	25	1080p25
1920 × 1080	16:9	Progressive	50	1080p50

<sup>a</sup> 50 Hz field rate. Note that other frame rates can be used. The most common are 24, 25, 30, 50, and 60 fps.

**Table 4.10** Basic HDTV image sizes in National Television System Committee countries

Size	Aspect ratio	Scan	Frame rate (fps/Hz)	Label
1280 × 720	16:9	Progressive	60	720p60
1920 × 1080	16:9	Interlaced	30 <sup>a</sup>	1080i30
1920 × 1080	16:9	Progressive	30	1080p30
1920 × 1080	16:9	Progressive	60	1080p60

<sup>a</sup> 60 Hz field rate. Note that other frame rates can be used. The most common are 24, 25, 30, 50, and 60 fps.

### 4.6.6 Ultra-HD resolutions

Developed in Japan and then standardized by the International Telecommunication Union (ITU) in 2012 and the SMPTE in 2013, Ultra-HD television (UHDTV) has two digital video formats:

Ultra-HD 2160p is by most referred to as 4K, which refers to the horizontal resolution of approximately 4000 pixels.

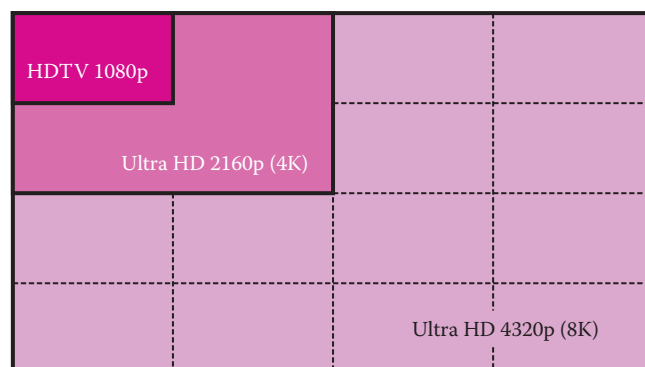
- 3840 × 2160 pixel resolution
- 8.3 megapixels—four times more pixels than HDTV 1080p (see Figure 4.50)
- Extrahigh color fidelity (color spaces are defined by ITU-R BT.2020)
- 16:9 aspect ratio
- Progressive scan
- Main compression standard being H.265, although H.264 is temporarily used
- Up to 100/120 fps
- Also known as UHDTV 2160p, UHD-1, 4K, or quadHD, although the term 4K was originally used for the cinema resolution of 4096 × 2160 and aspect ratio of 21:9

Ultra-HD 4320p is by most referred to as 8K, which refers to the horizontal resolution of approximately 8000 pixels.

- 7680 × 4320 pixel resolution
- 33.2 megapixels—sixteen times more pixels than HDTV 1080p (see Figure 4.50)
- Extrahigh color fidelity (color spaces are defined by ITU-R BT.2020)
- 16:9 aspect ratio
- Progressive scan
- Main compression standard being H.265, although H.264 is used temporarily
- Up to 100/120 fps
- Also known as UHDTV 4320p, UHD-2, 8K, Full Ultra-HD, or Super Hi-Vision

Figure 4.50 shows a comparison between HDTV 1080p and the two Ultra-HD resolutions.

As in the HDTV standards, the “p” indicates progressive scan. Among consumers in the home theater industry, the Ultra-HD promise of significantly better color fidelity has generated a lot of interest along with the advantages of higher resolution. In the video surveillance industry, color also makes a difference, but the unmatched resolution means exceptional digital zoom capabilities and extended field of view. These superpowers make Ultra-HD cameras ideal for capturing facial features and other fine details. Like HDTV, Ultra-HD is a standard that guarantees a certain video quality, whereas megapixel only is a statement of the camera’s resolution. However, Ultra-HD technology is still expensive. It demands high-quality lenses, a lot of pixels, and big sensors that can capture enough light, as well as significant expansions in bandwidth and storage capacity.



**Figure 4.50** Ultra-HD 4320p (8K) has 16× the resolution of HDTV 1080p. Ultra-HD 2160p (4K) has 4× the resolution of HDTV 1080p.

### 4.6.7 Aspect ratios

Aspect ratio is the ratio of the image's width to its height. Megapixel, multi-megapixel, HDTV, and Ultra-HD resolutions allow a greater flexibility in the choice of aspect ratio (Figure 4.51). An old-fashioned TV monitor, which traditionally is used to view analog surveillance video, displays an image with an aspect ratio of 4:3. Network video can offer the same ratio but also other ratios such as 16:9.

The advantage of a 16:9 aspect ratio is that unimportant details usually located in the upper and lower parts of the scene are not part of the image. Therefore, bandwidth and storage requirements can be reduced.

The aspect ratio 9:16 is also popular in video surveillance. It is as simple as a 16:9 image that has been rotated 90°. In a 9:16 aspect ratio, the full view of the camera can be utilized to overlook narrow scenes such as retail store aisles, highways and freeways, school hallways, and train station platforms (see Figure 4.52). This aspect ratio is sometimes referred to as corridor format.



**Figure 4.51** Different aspect ratios.





**Figure 4.52** Corridor format is often used in retail environments.

## 4.7 BEST PRACTICES

- *Select the right camera for the lighting situation:* A camera might deliver reasonable image quality in bright light conditions, but learn the challenges of the surveillance area to meet them in the best possible way: Will the camera be used in indoor, low-light environments, or will it need to cover both daylight and nighttime scenes? If possible, avoid backlight situations and extreme dynamic ranges altogether. If not possible, ask yourself: Are there curtains or blinds that can be closed? Can the camera be moved to a better position? Can light be redirected or frontal lights added? Can the sky be cut out of the image (large portions of bright skies increase the dynamic range)? What types of WDR does the camera support?
- *Select the right camera for the surveillance needs:* Different camera types and resolutions have different strengths. If the main objective is to get an overview of a large area or to focus on several smaller areas within a larger area, a multi-megapixel camera is probably a good choice. Is the surveillance area wide or narrow? Multi-megapixel cameras and lenses with a wide field of view produce great overview images of large areas. Select a camera that offers a suitable aspect ratio so that as much as possible the image fulfills its purpose.
- *Consider the lens:* A high-quality lens can deliver better images. Only fixed cameras typically have interchangeable lenses, and some lenses feature auto-iris control that improves the dynamic range. With remote focus and zoom, the user can adjust the focus and zoom through their computer. This makes installation and maintenance easier, faster, and cheaper. Is a fixed lens sufficient, or could the required field of view change over time? If so, a varifocal lens is probably a smarter choice. Often, the local authorities can make suggestion as to which is the best position to optimize the view. If the possibility to track and zoom in on subjects is required, consider a megapixel camera with digital PTZ, a varifocal lens, or even a camera with full mechanical PTZ capabilities. Make sure that the lens fits the sensor, that the field of view is appropriate, and that the focal length and aperture give the right depth of field for the camera-to-object distance.

- *Use image enhancement techniques based on requirements:* If used correctly and if the surveillance situation requires it, backlight compensation, WDR, and noise filters, as well as white balancing, sharpening, and contrast enhancement, can significantly increase image quality. Different techniques serve different purposes. To end up choosing the right settings for the right scene, play with settings and learn how the enhancements affect the images.
- *Look for progressive scan:* Progressive scan makes it possible to identify details, such as license plates, of moving objects and subjects in frozen frames. The images do not suffer from the saw effect that hampers interlaced video technologies. Nowadays, most surveillance cameras have progressive scan.
- *Understand file size and bandwidth requirements:* Network cameras use image compression. There is a trade-off between high-quality images and compressed images, which require much less bandwidth. Learn how motion, changes in lighting, and other parameters affect the scene and the bandwidth consumption. Different compression settings have different types of allowances for controlling bitrate and frame rate. Learn how they work and what allowances the typical scenario for a specific camera can have to maintain the optimal rates.

## CHAPTER 5

### Thermal cameras

Thermal network cameras create images based on the principles of infrared (IR) radiation. All objects and organisms generate some amount of heat. Heat is a form of light that is invisible to the eye and is also known as thermal IR radiation. The thermal camera serves as a heat sensor that detects temperature differences between objects and the scene itself. Rather than a picture of light, the thermal image is a visual capture of heat. The more heat an object emits, the brighter it appears in a thermal image.

Thermal images give the most information when there are significant temperature differences in a scene. Images are generally produced in grayscale where dark areas indicate colder temperatures and light areas warmer ones. Color palettes can be added to enhance different shades in the image.

Thermal cameras can detect people, objects, and incidents in low-light environments, complete darkness, or other challenging conditions such as smoke-filled and dusty environments. Thermal imaging has proved to be a lifesaver in emergency situations. See Figure 5.1 for illustrations of thermal camera detection.

Because thermal imaging does not provide sufficient information for identification, its primary use is to detect irregularities and suspicious activities. Thermal cameras quickly and accurately detect any incidents occurring in their field of view. They are robust and cannot be blinded by strong lights or put out of order using laser pointers. These qualities make them a great choice for first line of protection. Upon detecting an incident, they can immediately trigger further action, dramatically enhancing the effectiveness of a surveillance system.

Thermal cameras are perfect for perimeter or area protection. They are a powerful and cost-effective alternative to radio-frequency intruder detection, fences, and floodlights. Since thermal cameras do not need ambient light to produce images, they provide discreet surveillance in complete darkness. And in situations where some light is still needed for identification purposes, they can reduce the need for excess illumination. Thermal cameras can also improve security in off-limit spaces, for example, in transportation areas such as tunnels, railway tracks, and bridges.

Indoor uses include building security and emergency management. Thermal cameras can detect humans inside a building after business hours or in emergency situations, for example, when rooms fill up with smoke. High-security buildings, nuclear power plants, correctional facilities, airports, pipelines, and sensitive sections of railways also benefit from thermal camera surveillance.

Dual cameras offer a combination of conventional and thermal camera technology. They can provide a very wide range of detection and surveillance and are ideal for mission-critical applications where 24-hour monitoring is required (see Figure 5.2).



**Figure 5.1** Complete darkness, haze, smoke, rain, snow, and even bright and blinding lights—a thermal network camera is still able to detect people and objects.



**Figure 5.2** An example of a dual camera that combines conventional and thermal imaging capabilities.

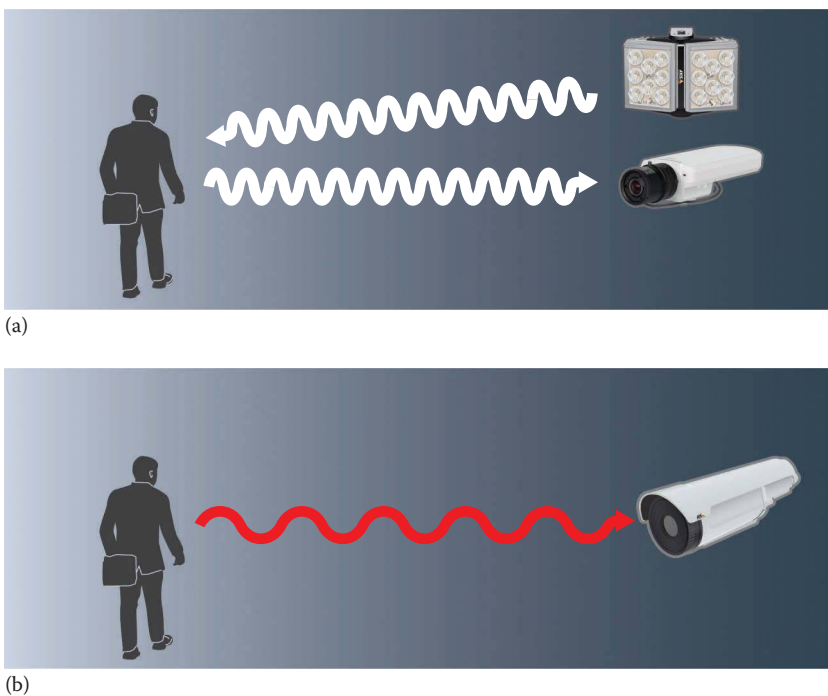
Thermal imaging technology is not new. But until recently, costs were prohibitive. This made practical applications outside the military, law enforcement, and high-security locations rare. The change is a result of new sensors. Streamlining sensor production and lens material improvements are driving volumes and making prices more reasonable. Today, thermal imaging is applied in various lines of business, such as the aircraft industry, shipping industry, and other critical infrastructure. The technology is also used in public services like firefighting and law enforcement. Lately, it has even appeared in consumer products, such as mobile devices and luxury cars.

This chapter gives an overview of the basic principles of thermal imaging, along with an overview of the components in a thermal camera. Calculation of detection range, integration with intelligent video analytics, and export regulations are other topics covered here.

## 5.1 HOW THERMAL IMAGING WORKS

An image from a conventional surveillance camera requires that light is reflected off of the object that is being photographed. Light is an electromagnetic radiation, which is interpreted into an image by light-sensitive silicon material in an image sensor. A thermal camera also collects electromagnetic radiation to create an image (see Figure 5.3).

However, while a conventional camera works in the range of visible light, with radiation wavelengths between approximately 400 and 700 nm (or 0.4–0.7  $\mu\text{m}$ ), a thermal camera is designed to detect radiation with greater wavelengths, up to around 14,000 nm (or 14  $\mu\text{m}$ ). Radiation in this part of the electromagnetic spectrum is referred to as IR, which in turn can be divided into several subgroups.



**Figure 5.3** The difference between conventional image capture and thermal image capture: How a regular camera detects light reflected off an object (a). How a thermal camera detects thermal radiation, which is emitted by all objects having a temperature above the absolute zero temperature, 0 K (−273°C or −459°F) (b).

### 5.1.1 Electromagnetic spectrum

The parts of the electromagnetic spectrum that are beyond visible light, the IR wave band, are often divided into the following subregions:

- Short-wave infrared, 1.4–3  $\mu\text{m}$
- Mid-wave infrared (MWIR), 3–8  $\mu\text{m}$
- Long-wave infrared (LWIR), 8–15  $\mu\text{m}$
- Far-wave infrared (FWIR), 15–1000  $\mu\text{m}$  (or 1 mm)

Earth scientists define the LWIR wave band as thermal IR, but in the thermal imaging industry, the MWIR wave band is also commonly referred to as thermal. However, the 5–8  $\mu\text{m}$  part of the MWIR wave band is virtually unusable for thermal imaging purposes because of the high spectral absorption of the atmosphere in this range. The thermal imaging industry often divides the electromagnetic spectrum based on the response of various IR detectors. Very long-wave infrared is added between LWIR and FWIR, and even the boundaries between the other ranges are slightly different (see Figure 5.4).

Microwaves have a wavelength above 1 mm. At the far end of the spectrum are radio waves, with a wavelength of 1 m or more. In the other end of the spectrum, wavelengths shorter than those of visible light are successively referred to as ultraviolet, x-rays, and gamma rays.

### 5.1.2 Near-infrared imaging

Near-infrared (NIR) imaging is not the same as thermal imaging. NIR light has a wavelength of about 0.7–1.4  $\mu\text{m}$ , which is just beyond what the human eye can see. Camera sensors can be built to detect and make use of this type of radiation. These day-and-night cameras, sometimes called IR cameras, use an IR-cut filter during the day. The IR-cut filter removes the IR light so that colors, as they are seen by the human eye, are not distorted.

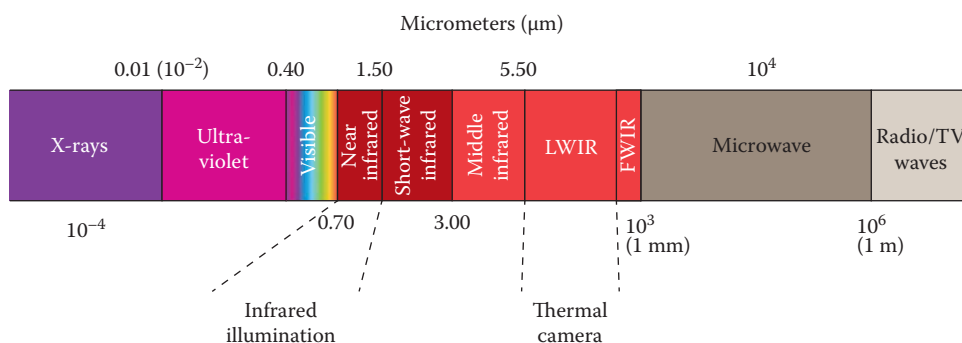
Figure 5.5 illustrates the differences in the images produced by conventional and thermal cameras.

When the camera is in night mode, the IR-cut filter is removed. To accommodate the human eye's inability to see IR light, the camera displays the image in grayscale. NIR imaging still requires some kind of light source—either natural, such as moonlight, or artificial light, such as streetlights or a dedicated IR light source. For more information about day-and-night cameras, see Chapter 3.

### 5.1.3 Using thermal radiation to create images

As explained earlier, what the human eye sees as images is in fact the light reflected by different objects.

No light means no reflection, and the eye is “blind” under such circumstances. Thermal images, on the other hand, are not dependent on visible light. Instead, images are created by operating in



**Figure 5.4** Different camera technologies work in different ranges of the spectrum of light.





**Figure 5.5** Infrared (IR) cameras switch to black-and-white mode to deliver better images in low-light scenes (a). Thermal cameras use differences in heat to provide detection in completely dark scenes and harsh environments (b). Even though the resolution is lower, the thermal camera can detect a person in situations where the IR camera cannot.

the thermal IR spectrum. This works perfectly well even in total darkness, since the ambient light level does not matter.

What makes thermal imaging possible is that all objects—organic and inorganic—emit a certain amount of thermal IR radiation as a function of their temperature. This is true for all objects with a temperature above absolute 0, or 0 K ( $-273^{\circ}\text{C}$  or  $-459^{\circ}\text{F}$ ). That means that even very cold objects, such as ice or an outdoor steel post in winter, emit thermal radiation.

The ability to emit absorbed energy is called emissivity. Depending on their properties, all materials have more or less emissivity ( $e$ ). The emissivity is measured on a scale from 0 to 1. The maximum value, 1, only applies for a theoretical object called a blackbody. Generally speaking, the duller and blacker a material, the more radiation it emits and the closer its value is to 1. Conversely, a more reflective material has a lower  $e$  value. For example, highly polished silver and brass have an emissivity of about 0.02 and 0.03, respectively. Iron has an emissivity of 0.14–0.035 if polished, but 0.61 if it has rusted red. Regular glass, which effectively blocks thermal radiation, has an emissivity of 0.92.

An object's thermal radiation is also dependent on its temperature. The hotter the object is, the more thermal radiation it emits. Although we humans cannot always see it, we can sense it. For example, we can feel the heat when we enter a sauna or step out on a hot tarmac. Objects with a high enough temperature, such as burning wood or melting metal, also radiate visible light that indicates the temperature of its surface.

The camera's sensitivity can be defined as its capability to distinguish between temperature differentials. The greater the temperature difference in a scene, the clearer the thermal images will be. But the contrasts in a thermal image also depend on the emissivity of its objects.

## 5.2 COMPONENTS OF A THERMAL CAMERA

At first glance, a conventional network camera and a thermal network camera might seem identical. Many things are indeed similar, such as the compression and the networking features with Power over Ethernet. Just like with conventional cameras, different form factors are available for use in different environments and situations: fixed indoor or outdoor cameras, dual conventional and thermal cameras, and cameras on pan-tilt heads or with zoom options. However, the two things that differ substantially are the lens and the sensor. This section explains these differences.

## 5.2.1 Sensors

The sensor in a thermal camera is an array of thousands of detectors that are sensitive to thermal IR radiation. The sensor detects, records, and then converts the thermal IR information into electrical signals. This is what makes the video image. Detectors used for thermal imaging can be broadly divided into two types: cooled and uncooled IR sensors.

Uncooled IR image sensors are smaller and built with fewer moving parts, which makes them less expensive than their cooled counterparts. Unlike cameras with uncooled sensors, cameras with cooled sensors generally need to be serviced, although newer designs based on Sterling Motors need less service. Also, the cooling medium must be refilled every 8,000–10,000 hours.

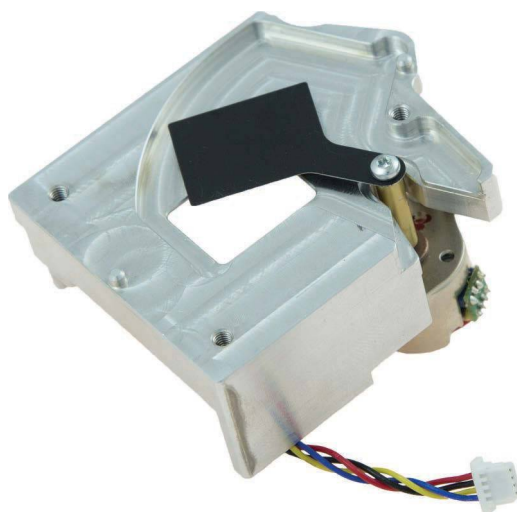
The individual elements in an uncooled sensor respond in different ways to the incoming IR radiation, which results in a “drift” in individual pixel values. To remedy this, the sensor performs nonuniformity correction. A mechanical shutter blocks the sensor and gives it a standard temperature target, against which every pixel is corrected. This process occurs at regular intervals or when a specific temperature change takes place. A typical shutter for a thermal camera is shown in Figure 5.6.

### 5.2.1.1 Cooled sensors

Cooled IR sensors are usually contained in a vacuum-sealed case and are cooled to temperatures as low as 60 K to 100 K (approximately from  $-210^{\circ}\text{C}$  to  $-170^{\circ}\text{C}$  or from  $-346^{\circ}\text{F}$  to  $-274^{\circ}\text{F}$ ), depending on the type and level of performance desired. These extremely low temperatures are accomplished with so-called cryogenic coolers. Cooling is needed to reduce thermally induced noise. Otherwise, at higher temperatures the sensors risk being flooded or “blinded” by their own thermal radiation. This equipment makes the detectors relatively bulky, expensive, and rather energy consuming.

Although cooled sensor technology is both expensive and high maintenance, it has benefits. These detectors work in the midwave spectral band (MWIR), which provides better spatial resolution because the wavelengths are much shorter and deliver higher contrast than in the long-wave band. Hence, cooled detectors can distinguish smaller temperature differences and produce crisp, high-resolution images. See Figure 5.7 for an example of a cooled sensor.

Another advantage with cooled sensors is that the greater sensitivity also allows the use of lenses with high f-numbers (or f-stops). Consequently, cooled detectors are a better choice for long-range detection, that is, 5–16 km (3–10 miles). Figure 5.7 shows a cooled sensor.



**Figure 5.6** A mechanical shutter for a thermal camera. The shutter is used for temperature calibration of the sensor.



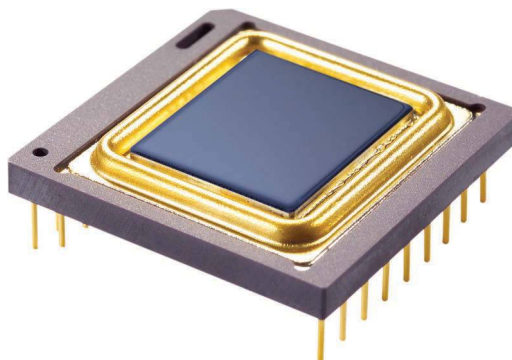
**Figure 5.7** A cooled thermal sensor with cryogenic cooling unit. (Image courtesy of Sofradir—Scorpio LW, Palaiseu, France.)

#### 5.2.1.2 Uncooled sensors

The sensor in an uncooled thermal camera is not dependent on cryogenic cooling. The uncooled IR sensor is stabilized at or close to the ambient temperature, using less complicated temperature control elements or no temperature control at all. Sensors of this kind operate in the LWIR band.

Uncooled sensors can be based on a variety of materials that all offer unique benefits. A common design is based on microbolometer technology. Typically, this is a tiny resistor (or thermistor) with highly temperature-dependent properties on a silicon element that is thermally insulated. The resistor is made of vanadium oxide (VOx) or amorphous silicon ( $\alpha$ -Si). When the thermal IR radiation hits the material, the electrical resistance changes.

Another kind of microbolometer is based on ferroelectric technology. Here, small changes in the material's temperature create large changes in electrical polarization. Ferroelectric microbolometers are made of barium strontium titanate (BST). Figure 5.8 shows an uncooled sensor.



**Figure 5.8** An uncooled infrared sensor, ULIS  $640 \times 480$  17  $\mu\text{m}$ . (Image courtesy of ULIS, Veurey-Voroize, France.)

Changes in scene temperature cause changes in the bolometer, which are then converted into electrical signals and processed into an image. The camera's sensitivity to thermal radiation, which determines its ability to distinguish different temperature differences in a scene, can be expressed as its noise equivalent temperature difference (NETD) value. Most thermal network cameras have an NETD value of 50–100 mK, though there are newer generations of bolometers that have an NETD as low as 20 mK.

## 5.2.2 Sensor resolutions

Resolutions are generally much lower for thermal cameras than for conventional network cameras. This is mostly due to the more expensive sensor technology involved in thermal imaging. The pixels are larger, which affects the sensor size and the cost of materials and production. Currently, typical resolutions for thermal cameras range from  $160 \times 128$  to high resolutions of  $640 \times 480$  (VGA), though even higher resolutions are available.

In visual observation, the larger image sensor delivers higher resolution and better image quality (see Figure 5.9).

## 5.2.3 Lenses for thermal cameras

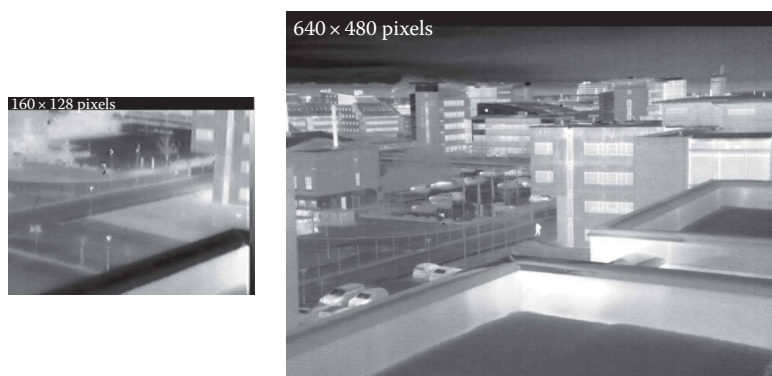
Because regular glass blocks thermal radiation, manufacturers cannot use regular glass-based optics and lenses in thermal cameras. Currently, germanium is the most commonly used material for thermal camera optics. This very expensive metalloid, which is chemically similar to tin and silicon, blocks visible light while letting through the IR light (Figure 5.10).

Not all lenses are pure germanium. For example, some are made of a germanium-based material called chalcogenide glass, which allows a wider spectrum of IR light to pass through. Like with most materials, there are benefits and disadvantages. Chalcogenide glass contains cheaper materials and is moldable. However, the master mold requires a significant initial investment that can only be justified at larger quantities. Thermal cameras use different lens mounts than conventional network cameras. The mount needs to be wider to fit the sensor, which is typically larger than a conventional sensor. A TA-lens has an  $M34 \times 0.5$  screw mount, allowing for sensors as large as 13 mm in diameter. Figure 5.11 shows examples of lenses for thermal cameras.

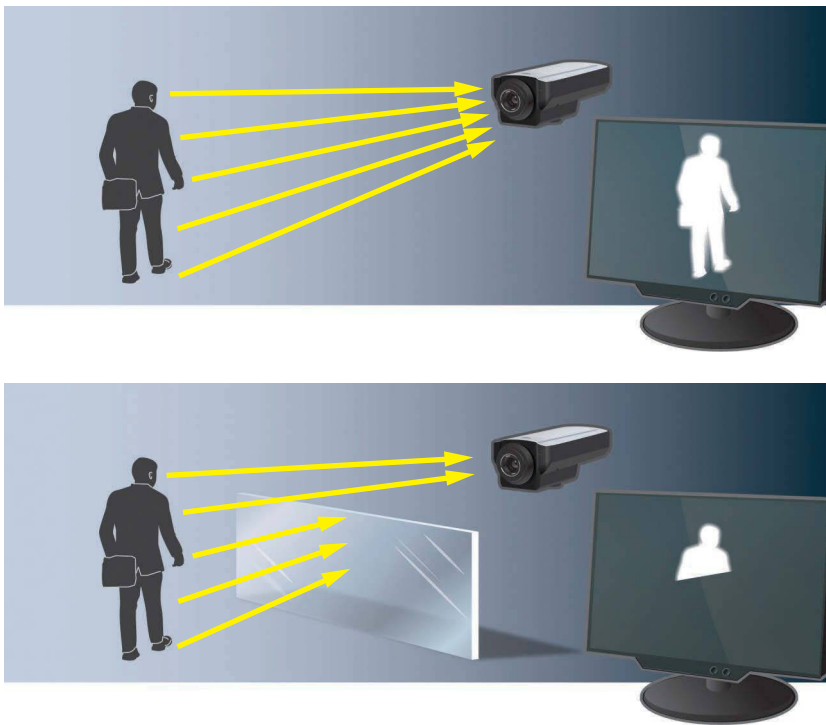
### 5.2.3.1 Calculation of focal length

The focal length of a lens is defined as the distance between the entrance lens (or a specific point in a complex lens assembly) and the point where all the light rays converge to a point, which normally is the camera's image sensor.

Like conventional lenses, thermal lenses come in different focal lengths, which are usually specified in millimeters (e.g., 35 mm). A longer focal length results in a narrower field of view. The field of view depends on the focal length and the diameter of the sensor. Because varifocal and zoom lenses



**Figure 5.9** Effect of sensor size on resolution and image quality.



**Figure 5.10** Glass preventing thermal imaging.



**Figure 5.11** Examples of lenses for thermal cameras.

need more lens material (more germanium), they are often too expensive to justify production and purchase. This is why fixed lenses are more common. Table 5.1 shows the relationship between focal length and field of view.

While focal length is specified for lenses, we sometimes need to know which lens to use for a specific application. Nomographs are used to determine the relationship of the focal length of the lens, the number of pixels across the object, and the range. See Section 5.4 for more information.

### 5.2.4 Thermal enclosures

The main environmental threats to a network video product, particularly one that is installed outdoors, are cold, heat, water, dust, snow, and insects. Sometimes, this means that a thermal camera needs a protective enclosure. Naturally, the same transparency to IR light applies to housings,

**Table 5.1** Focal length and field of view

Focal length (mm)	Horizontal field of view (°)	Sensor size (pixels)
7	55	384 × 288
10	51	384 × 288
	57	640 × 480
13	17	160 × 128
	28	384 × 288
19	28	384 × 288
	32	640 × 480
35	16	384 × 288
	18	640 × 480
	10.7	384 × 288
60	9	384 × 288
	10	640 × 480
	6.2	384 × 288



**Figure 5.12** A typical enclosure of a thermal camera (a) is similar to that of a conventional surveillance camera (b) At first glance, they look very similar except for the material used in the window in front of the camera. But their imaging capabilities are very different.

making it impossible to use glass in windows® of standard housings. Like lenses, housings must be specially adapted for thermal cameras. Therefore, germanium is used instead.

For more information about environmental conditions, enclosures and operating range, see Chapter 17 about system design.

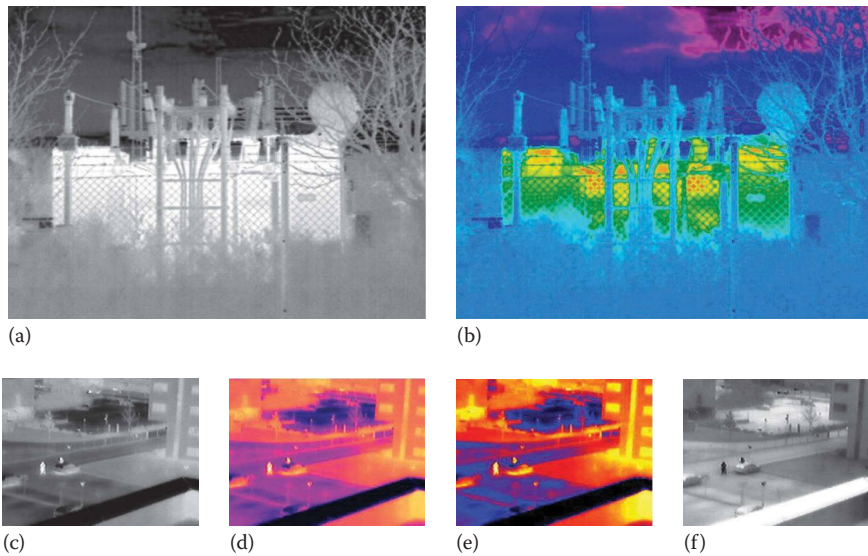
Figure 5.12 shows an outdoor-ready thermal camera and an outdoor-ready conventional camera.

### 5.3 PRESENTATION OF THERMAL IMAGES

The most common presentation of thermal images is in black, white, and gray. The different temperature values are then translated into 256 grayscale values. The most common presentation (or palette) is *white-hot*, in which heat sources appear white against lower-temperature gray and black backgrounds. In some cases, *black-hot*, where sources of heat appear black, may be easier to use. Most cameras can switch between palettes.

Thermal images are sometimes associated with bright, intense colors, which may seem a little odd considering that the camera works outside the spectrum of visible light. Since the human eye is better at distinguishing different shades of color than different shades of gray, adding color sometimes makes it easier to see differences in thermal images. These so-called pseudocolors are created digitally.



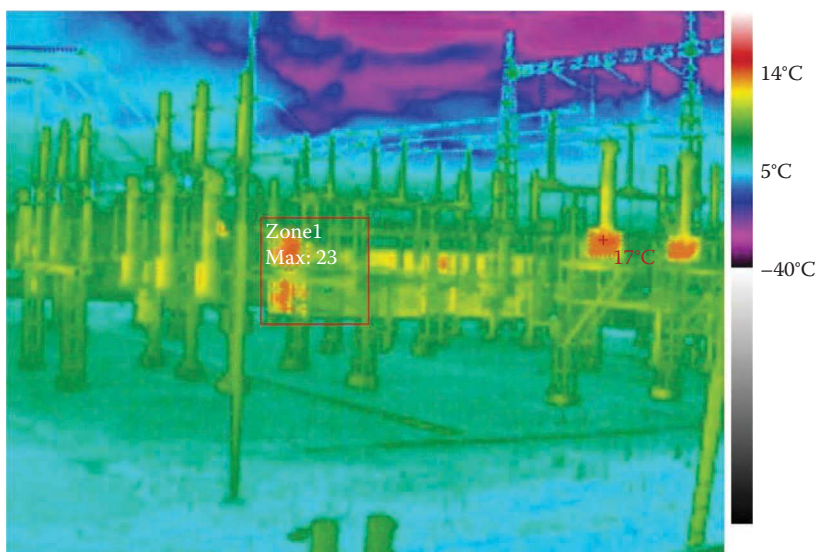


**Figure 5.13** Thermal images with five different palettes: white-hot (a, c), black-hot (f), and different types of pseudocolor (b, d, and e).

Each color or nuance represents a different temperature, usually ranging from white and red for higher temperatures to green, blue, and violet for colder ones. Figure 5.13 compares different kinds of thermal image presentation.

### 5.3.1 Temperature alarm cameras

Temperature alarm cameras are based on thermal imaging and use the same sensor technology as thermal cameras. They can be used for remote temperature monitoring and include the possibility to set temperature alarms. There are two basic types of temperature alarms. One is triggered when the temperature goes above or below the set temperature limit but also if the temperature changes too quickly. The other is a spot temperature alarm, where the camera measures the temperature of a specific area in the image. Isothermal palettes highlight the critical temperatures so they stand out from the rest of the scene. Figure 5.14 shows an example of an isothermal palette applied on a food processing plant.



**Figure 5.14** The isothermal palette makes it possible to highlight the temperature span and easily see if a surface reaches a hazardous temperature.

5.4 DETERMINING DETECTION RANGE


The resolution required to detect an object is stated in pixels and is usually determined by means of Johnson’s criteria. John Johnson, a U.S. military scientist, developed this method for predicting the performance of sensor systems during the 1950s. Johnson measured the ability of observers to identify scale model targets under various conditions and came up with criteria for the minimum required resolution. These criteria give a 50% probability of an observer discriminating (detecting, recognizing, or identifying) an object at the specified level. This object can be a person, typically defined with a critical dimension of 0.75 m (2.46 ft), or a vehicle, typically defined with a critical dimension of 2.3 m (7.55 ft).

Johnson’s criteria assume ideal conditions. In reality, the weather conditions at the site affect the thermal radiation emitted from the object and decrease the effective detection range. According to Johnson’s definition, the detection range in Table 5.2 ideally requires a temperature difference of 2°C (3.6°F) between the targeted object and the background. However, certain weather conditions, such as rain, snow, and fog, attenuate the radiation emitted from the object. This is because the heat that radiates from the object scatters when it hits particles in the air. The camera should always be evaluated in the intended location and environment to avoid performance and reliability problems. The levels of Johnson’s criteria used for thermal network cameras are as follows:

- *Detection*: At least 1.5 pixels are needed for the observer to see that an object is present.
- *Recognition*: At least 6 pixels are needed for the observer to distinguish the object, for example, a person in front of a fence.
- *Identification*: At least 12 pixels are needed for the observer to both distinguish an object and its characteristics, such as a person holding a crowbar.

Johnson’s criteria were developed under the assumption that visible information was processed by a human observer. If the information instead is processed by an application algorithm, there will be specific requirements on the number of pixels needed on the target for reliable operation.

Table 5.2 Example of a thermal camera range chart

						
	Focal length mm	Horizontal field of view Degrees	Human: 1.8 × 0.5 m (5 ft 11 in. × 1 ft 8 in.) Critical dimension: 0.75 m (2 ft 6 in.)		Vehicle: 1.4 × 4.0 m (4 ft 7 in. × 13 ft 2 in.) Critical dimension: 2.3 m (7 ft 7 in.)	
			Meters	Yards	Meters	Yards
<i>Detection</i> (1.5 pixels on target)	10	51	220	241	660	722
	19	28	390	427	1200	1312
An observer can see an object.	35	16	700	766	2200	2405
	60	9	1200	1312	3700	4046
<i>Recognition</i> (6 pixels on target)	10	51	55	60	170	186
	19	28	100	109	300	328
An observer can distinguish an object.	35	16	175	191	550	601
	60	9	300	330	920	1006
<i>Identification</i> (12 pixels on target)	10	51	25	37	85	93
	19	28	50	55	150	164
An observer can distinguish a specific object.	35	16	90	98	270	298
	60	9	150	165	460	503

This sample of a thermal camera range chart specifies at which distances humans and vehicles can be detected, recognized, and identified.

All video analytics algorithms need to work with a certain number of pixels. The exact number may vary. But as a rule of thumb, at least 6 pixels across the object are required, which is the same level as Johnson's criteria for recognition. Even if a human observer would be able to detect the object, to work as intended, the application algorithm often needs a larger amount of pixels at a given detection range.

### 5.4.1 Nomograph

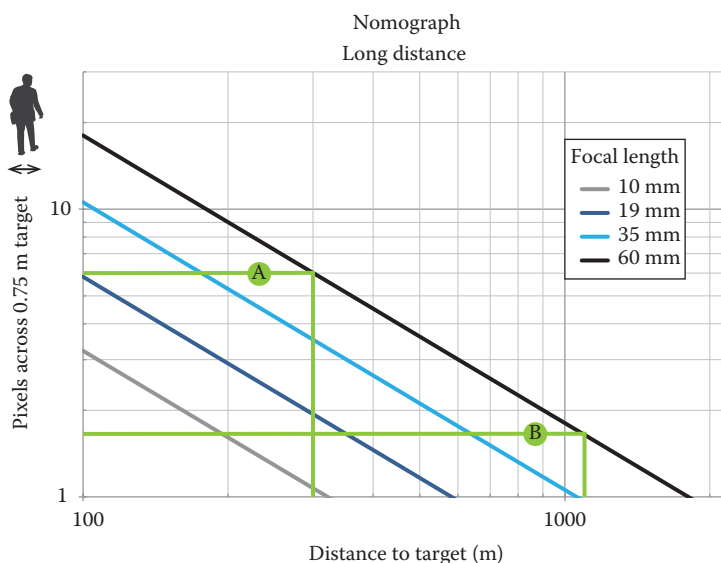
A nomograph is used to find out the available number of pixels at a given range. It is a two-dimensional diagram that shows the relation between the focal length of the lens, the number of pixels across the object, and the range. For example, if the number of pixels required and the distance at which an object needs to be recognized are known, it is possible to calculate which lens or camera to use. Equally, if the camera and the number of pixels required are known, the distance at which the camera can be used to detect an object is indicated by the nomograph (see Figure 5.15).

Here is an example: a thermal camera with a 60 mm lens points at a person with a critical dimension of 0.75 m (2 ft 5 ½ in.). The nomograph in Figure 5.14 shows that the object is recognizable at 300 m (328 yd) and 6 pixels across the object (A). If only detection is required, the range is 1200 m (1312 yd) and 1.5 pixels across the object (B).

### 5.4.2 Environmental considerations

Remember that Johnson's criteria are only valid in ideal conditions. The weather conditions on-site affect the thermal radiation emitted from the object and decrease the effective detection range. The detection range used in the nomograph earlier ideally requires a temperature difference of 2°C between the targeted object and the background. This section further explains how environmental factors influence the thermal camera's performance.

Environmental factors that affect the thermal camera's ability to produce an image include weather conditions and the temperature difference between the object and its background. An object with almost the same temperature as the background, such as a body on a hot summer day, is harder to distinguish from its background than an object that is considerably hotter or colder, such as a car with a running engine on a cold winter day.



**Figure 5.15** An example of a nomograph.

**Table 5.3** Different environmental conditions cause different levels of attenuation

Weather conditions and attenuation				
Heavy rain	Light rain	Urban pollution	Dense fog	Fog
11 dB/km	4 dB/km	0.5 dB/km	80 dB/km	10 dB/km
17.6 dB/mile	6.4 dB/mile	0.8 dB/mile	128 dB/mile	16 dB/mile

The two most important environmental factors that affect the camera’s image of an object are absorption and scattering. They reduce the thermal radiation that reaches the camera and therefore shortens the distance at which the camera can detect an object. Scattering has a greater effect on the loss of thermal radiation than absorption.

Table 5.3 describes examples of attenuation in various environmental conditions. For example, on a foggy day the attenuation is 10 dB/km or 1 dB/100 m. For example, if the thermal camera is placed 300 m (980 ft) away, the attenuation for that distance totals 3 dB. A three-decibel attenuation means that the thermal camera receives 50% less thermal energy than if the conditions were optimal. The thermal energy reduction can be calculated with the following formula where  $a$  is attenuation as a negative decibel value:

$$\text{Thermal energy reduction} = 100\% \times 10^{(a/10)}$$

$$\text{Thermal energy reduction} = 100\% \times 10^{-3/10} = 50.1\%$$

#### 5.4.2.1 Absorption

Water vapor ( $H_2O$ ) and carbon dioxide ( $CO_2$ ) in the air are the primary causes of absorption. During absorption, the heat radiated from the object is absorbed by water vapor and carbon dioxide and loses some of its intensity before reaching the camera. Absorption usually affects the background more than the objects in the foreground.

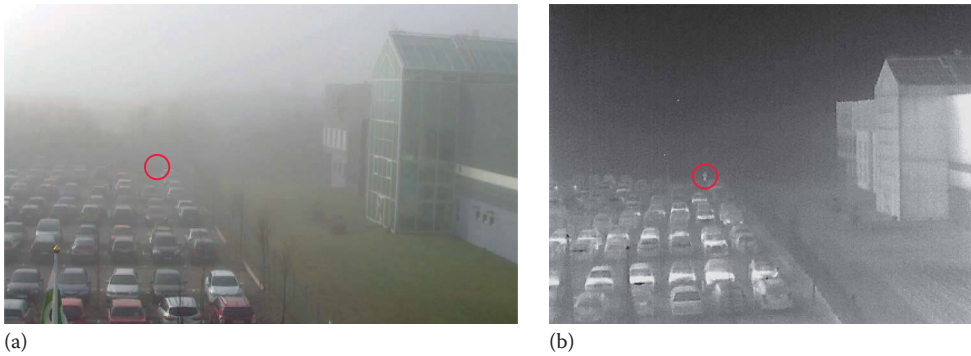
The water vapor content of the air affects image quality even in sunny and clear weather. In winter, if all other weather conditions are the same, the water vapor content of the air is lower than in summer. Because water vapor content is lower in winter, less thermal radiation is absorbed by the water molecules. More thermal radiation reaches the thermal network camera, and the result is an image with better quality than it would have on a summer day.

#### 5.4.2.2 Scattering

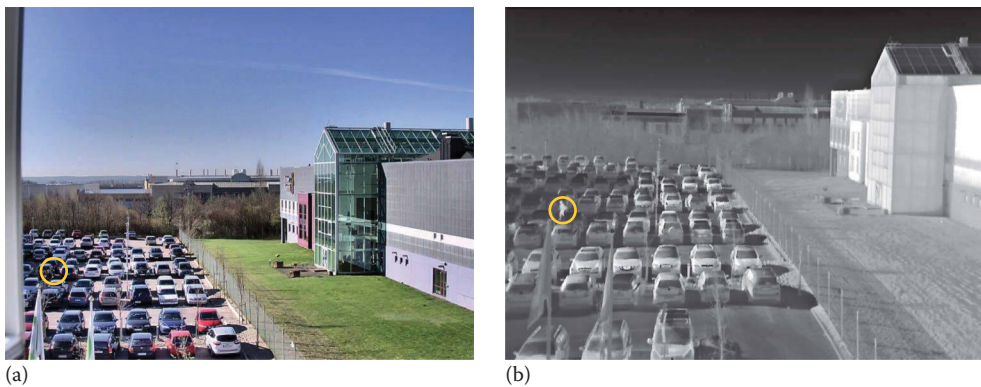
The thermal radiation the object emits is dispersed when it hits particles in the air. This is called scattering. The loss of radiation is directly related to the size and concentration of the particles, droplets, or crystals that constitute polluting, condensing, or precipitating conditions, such as smog, fog, rain, or snow.

Fog appears when water vapor in the air condenses into water droplets. The droplet sizes vary with different kinds of fog. In dense fog, the water droplets are bigger due to piling (or accretion), thus scattering thermal radiation more than light fog. Also, fog scatters thermal radiation to a larger extent than both smog and haze because of the greater size and concentration of its water droplets.

Like in this book, the effects of scattering are often mentioned in relation to thermal images. However, this does not necessarily mean that a conventional camera performs better in such conditions (Figures 5.16 and 5.17).



**Figure 5.16** These images were taken the same day by a conventional network camera (a) and a thermal network camera (b). Despite the difference in resolution ( $1280 \times 800$  pixels and  $384 \times 288$  pixels, respectively), the thermal camera provides better detection capabilities.



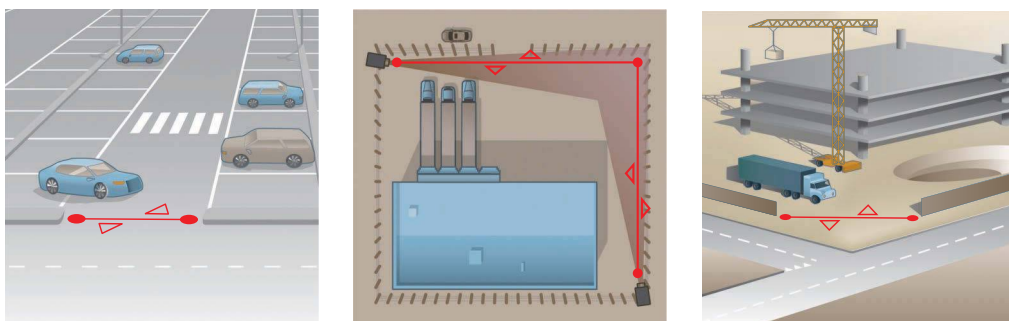
**Figure 5.17** These images were taken on a sunny day by a conventional network camera (a) and a thermal network camera (b). Even in broad daylight, the thermal image delivers contrast in areas where a conventional color image would not.

## 5.5 INTEGRATING THERMAL CAMERAS WITH INTELLIGENT VIDEO

A conventional network camera reacts to changes in the captured image and can, for example, be disturbed by shades and backlighting. A thermal network camera detects the thermal radiation from the object, which is a more static parameter compared to visual changes in an image. Therefore, a thermal camera is an especially reliable platform for integrating intelligent video applications (analytics) such as the following:

- *Motion detection:* When a thermal network camera has detected a moving object, it can be programmed to automatically perform or relay actions, such as sending an alarm to an operator, turning on floodlights, or triggering a regular camera to capture visual information about the incident. The camera only records during the actual incident, which means the recording space is minimized. This facilitates video analysis and saves valuable operator time.
- *Crossline detection or perimeter detection:* A virtual line can be placed in the image of the thermal network camera (see Figure 5.18). The virtual line acts as a tripwire. As in the motion detection example, the thermal network camera can trigger another camera if an object crosses the virtual tripwire.





**Figure 5.18** Crossline detection sets up a virtual line. When the line is crossed, the camera can be triggered to perform actions such as turning on floodlights and sirens or trigger another camera to zoom in on the scene.

When motion is detected or a virtual line is crossed, the camera can automatically trigger an alarm to the operator and at the same time trigger a PTZ camera to supply video to the operator. Based on the information, the operator can decide about the correct action to take.

Integrating thermal network cameras with intelligent video applications, also known as analytics, has many advantages. However, to get the optimum use of thermal network cameras, other things have to be considered than when using conventional network cameras. The definition of detection range, the number of pixels across the object, and the surrounding environment are all parameters that need to be considered. And they are of special importance when integrating with an intelligent video application. For more information, see Section 5.7 and Chapters 15 and 16.

## 5.6 EXPORT REGULATIONS FOR THERMAL TECHNOLOGIES

Technical aspects, legal considerations, and other hurdles present challenges when integrating thermal cameras into the conventional video surveillance market.

A special group of products and technologies that can be used both for military purposes and in commercial settings are called dual-use goods. Exports of such items are regulated in the international Wassenaar Arrangement from 1996, which, among other things, aims to promote transparency and greater responsibility in transfers of conventional arms as well as dual-use goods and technologies.

For a thermal camera to be freely exported, its maximum frame rate cannot exceed 9 frames per second (fps). Thermal cameras with a frame rate of up to 60 fps can be sold in EU countries, Norway, Switzerland, Canada, the United States, Japan, Australia, and New Zealand on the condition that the buyer is registered, can be traced, and has an export license.

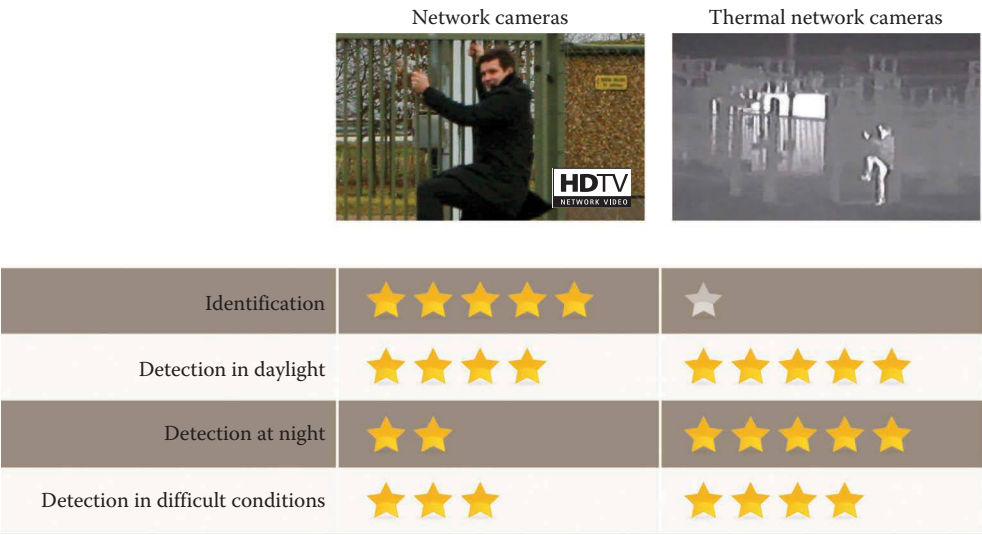
## 5.7 BEST PRACTICES

Thermal imaging is becoming an integral part of video surveillance systems, and it is becoming less expensive—putting thermal imaging within affordable reach for an ever-growing market. Thermal cameras can be an excellent complement in many situations where conventional cameras are inadequate. To see what can be the best application areas for thermal cameras, see Figure 5.19, which shows a feature comparison between conventional and thermal cameras.

Best practices for thermal cameras include the following:

- *Situations of total darkness:* Thermal cameras are unparalleled when light is completely absent. They can also be an option in areas that are very difficult to illuminate effectively, such as sea-fronts, harbors, or other large expanses of open water. Artificial light not only runs the risk of revealing where the cameras are placed, making them easier to avoid or vandalize, but can also project shadows that help intruders avoid detection.





**Figure 5.19** Comparison of features in conventional cameras and thermal cameras.

- *When detection is preferred over identification:* For privacy reasons, it is sometimes better to be able to detect people but not have the ability to identify them. Examples of such a scenario are a train platform or fenced in areas where you want to be able to detect when an object enters the area. This scenario is a perfect use case for thermal cameras.
- *When lighting is an issue:* Spotlights can blind as well as illuminate. Cameras that do not rely on light can be the preferred solution in many different traffic situations such as railway tunnels, airstrips, or ordinary streets. Thermal cameras cannot be blinded by bright lights or laser beams, nor do they rely on light to do their job.

A thermal camera can meet demanding surveillance requirements, but for the camera to do its job, it is essential to determine the correct lens and resolution for the application by using Johnson’s criteria.

The advantages of a thermal camera can be maximized when they are combined with intelligent video solutions that can help analyze and use the thermal images. For example, thermal cameras can be used to detect the presence of persons in dark areas outside a building, and an intelligent video system can warn security staff when this occurs. This means that personnel do not spend unnecessary time monitoring activity-less video streams.



# Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

## CHAPTER 6

# Video compression technologies

Analog video contains a tremendous amount of data, and when digitized and transmitted, these data can consume as much as 165 Mbit/s (megabits per second) of bandwidth.

It is not practical to transmit this amount of data over an IP-based network, and it is a challenge to store it in a cost-effective manner. Therefore, image and video compression techniques are used to reduce the bitrate. The goal is to drastically decrease the amount of data that must be sent but with as little negative effect as possible on the image and video quality. Depending on the purpose, different compression formats can be used. These formats are usually referred to as codecs. Codec is the abbreviated form of *compressor-decompressor* or *coding-decoding*.

This chapter introduces the basics of compression and the most common compression formats and the standardization organizations behind them, as well as more information about JPEG and H.264 compression, new trends, and best practices in the field.

## 6.1 BASICS OF COMPRESSION

The primary purpose of compression techniques is to reduce the amount of image data by discarding unnecessary information. As mentioned earlier, a digitized analog video sequence can comprise up to 165 Mbit/s of data. Those data could include large amounts of information that is unimportant for the purpose. For example, there is little surveillance benefit in a video where a big portion of the image is a white wall or blue sky.

All compression techniques are based on an understanding of how the human brain and eyes work together to form a complex visual system. An image can be optimally compressed, but if it is not viewed in the original size or rate, artifacts (block-like or blotchy effects in the image) may be visible. To effectively be able to pause or zoom in on a recorded image in a video sequence, recorded video should have a slightly larger bitrate (size) than what is needed for live viewing.

Some of the techniques commonly employed to reduce the size of images and video sequences include the following:

- *Quantization* reduces color nuances within an image.
- *Subsampling* reduces color resolution.
- *Transform coding followed by quantization* removes small invisible parts from the image.
- *Intraprediction* predicts how parts of an image will look based on adjacent parts in the same image.
- *Run-length* coding or prediction removes repeated pixel values.

- *Entropy coding* is a lossless data compression scheme (such as Huffman coding) that efficiently codes pixels.
- *Interprediction* compares adjacent images and removes unchanged details.

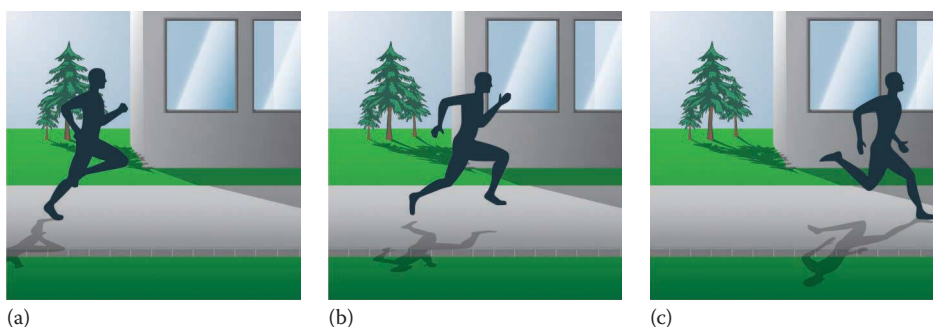
Some techniques are image-based compression techniques, also called intraframe compression, where only one frame is evaluated and compressed at a time. Others are video compression techniques, or interframe compression, where several adjacent frames are compared to reduce the amount of image data.

### 6.1.1 Image and video compression

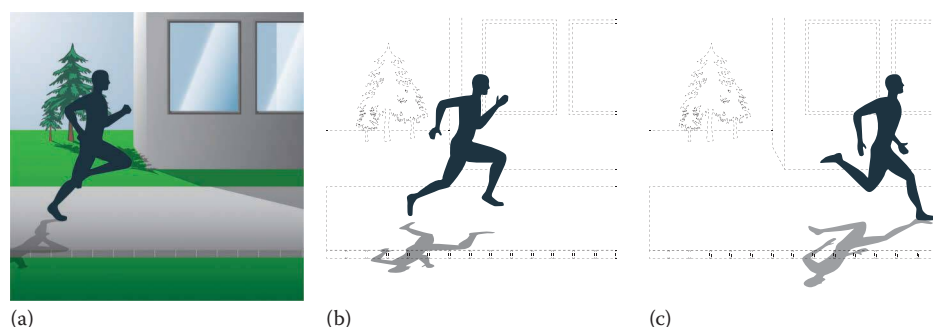
Consider the video sequence illustrated in Figure 6.1. The diagram shows a man running from left to right and a house and trees in the background. Image compression techniques code each image in the sequence as a separate, unique image.

Then consider Figure 6.2. When the same video sequence is encoded using video compression techniques, the static parts of the image are only included in the first frame. In the following two frames, only the areas in motion (the running person) are included. Because an encoded sequence that uses a video compression technique contains less information, less bandwidth and storage are required. When the encoded sequence is displayed, the images appear to the human eye just as it does in the original video sequence.

For more information about frame types, see Section 6.5.1.



**Figure 6.1** Motion JPEG encodes and sends the three images (a–c) in this sequence as separate, unique images (I-frames) with no dependencies on each other.



**Figure 6.2** Video codecs encode and send only the first image (I-frame) in its entirety (a). In the two P-frames shown here, references are made to the first picture for the static elements (the house) (b, c). Only moving parts (the running person) are coded using motion vectors. Therefore, the amount of sent and stored information is smaller.

### 6.1.2 Lossless and lossy compression

The two basic categories of compression are lossless and lossy compression. Lossless compression involves techniques that reconstruct the exact original data from the compressed data. However, while it might be advantageous to be able to retrieve complete data in some cases, these techniques do not provide significant data reduction. Graphics Interchange Format (GIF) is an example of a lossless image compression technique. Because of its limited compression abilities, GIF is not suitable for video surveillance.

Lossy compression, on the other hand, means that so much data will be discarded that the original information cannot be reconstructed when the video is decompressed. It is the absence of the discarded data that forms the artifacts mentioned earlier in this chapter.

### 6.1.3 Block transform

The compression algorithm can be pixel based, line based, or block based with selectable block size. A block is based on a number of pixels. For example, the JPEG format (named for the Joint Photographic Experts Group) uses  $8 \times 8$  blocks of 64 pixels. The block size is the smallest complete part that algorithms can use to make calculations. Many popular algorithms use block transform to sort the data before compression.

### 6.1.4 Prediction

Prediction is another image and video compression technique. It means forecasting the next pixels based on another nearby block of pixels. If prediction is performed well, an image with a blue sky can be encoded by transferring one blue pixel and instructions to use the same pixel to create a complete sky section in the image.

Prediction can also be extended to predict the next image once the first image is sent. Image-to-image prediction is used in video compression algorithms but not in image compression algorithms. This is the basic difference between the two types of algorithms. For more information about image-to-image prediction, see Section 6.5.1.

### 6.1.5 Latency

In compression, one or several mathematical algorithms have the job of removing image data. Similarly, to view a video file, algorithms are applied to interpret and display the data on a monitor. This process requires a certain amount of time, and the resulting delay is known as compression latency. Given the same processing power, the more advanced the compression algorithm, the higher the latency.

Today, the increasing computing power available in network cameras and PC servers makes latency less of a problem. However, the latency of a network and the processing power needed to avoid it must still be considered when designing a network video system.

In some contexts, such as the compression of studio movies, compression latency is irrelevant because the video is not viewed live. Low latency is essential in video surveillance applications where live video is monitored, especially when pan, tilt, and zoom (PTZ) cameras are used.

Normally video compression techniques have higher latency than image compression techniques. However, if a video compression technique such as H.264 is used, latency can be minimized by using a one-pass, network video-optimized encoder that avoids using double image references (B-frames) for prediction. For each B-frame that is added between P-frames, 40 milliseconds must be added to compensate for out-of-order encoding. For more details on frame types, see Section 6.5.1.

### 6.1.6 Jitter

Jitter is an artifact that causes parts of images to appear on the client monitor at the incorrect time. This can occur when video is delayed on the network or if the frame rate differs between the camera and the monitor. A video stream from the camera contains time codes, time stamps, which indicate when the image was captured. This information is transported to the client along with the video so that the viewing client can display the video correctly. The time stamps are also used to synchronize audio with its corresponding video playback.

The screen frequency must also be the same as the camera capture frequency. In Europe, a 50 or 100 Hz monitor must be used. In the United States, a 60 or 120 Hz monitor must be used.

### 6.1.7 Compression ratio

Compression ratio is defined as the ratio of the compressed bitrate to the uncompressed original bitrate. A 50% compression ratio means that 50% of the original data have been removed and the resulting video stream has only half the bitrate of the original.

With efficient compression techniques, significant reduction in bitrate can be achieved with little or no adverse effect on visual quality. The extent to which image modifications are perceptible depends on the amount of data that has been discarded. Often 50%–95% compression is achievable with no visible difference, and in some scenarios a compression ratio of more than 98% is possible.

## 6.2 COMPRESSION STANDARDS

Standards are important to ensure compatibility and interoperability. They are particularly relevant to video compression because video can be used for different purposes and sometimes must be viewable many years from the recording date.

In the mid-1990s, when storage was relatively expensive and standards for digital video compression were new, many video surveillance manufacturers developed proprietary video compression techniques. Today, most vendors use standardized compression techniques because they are equally good or better than the proprietary techniques. The increased use of standards means that end users can pick and choose from different vendors rather than being limited to a single supplier.

### 6.2.1 ITU and ISO

Two organizations are significant in the development of image and video compression standards: the International Telecommunication Union (ITU) and the International Organization for Standardization (ISO).

The ITU is not a formal standardization organization. The ITU stems from the telecommunications world and releases its documents as recommendations, for example, the ITU-R Recommendation BT.601 for digital video.

The ISO is a formal standardization organization that cooperates with the International Electrotechnical Commission (IEC) to develop standards within areas such as IT. The ISO is a general standardization organization, and the IEC is a standardization organization dealing with electronic and electrical standards. Often these two organizations are referred to as a single body: ISO/IEC.

However, given the ongoing convergence of communications and media, the organizations and their members have experienced an increasing overlap in their standardization efforts.

### 6.2.2 History of compression formats

The two basic compression standards are JPEG and MPEG (named for the Moving Picture Experts Group). Both are international standards set by the ISO, IEC, and contributors from the United States, Europe, and Japan, among others. JPEG and MPEG are also recommended by the ITU, which



has helped to establish them further as global standards for digital still image and video encoding. The Video Coding Experts Group (VCEG) is a subgroup within the ITU that has developed standards such as the H.261 and H.263 recommendations for videoconferencing over telephone lines.

The development of JPEG and MPEG standards began in the mid-1980s when the Joint Photographic Experts Group was formed. Seeking to create a standard for color image compression, the group's first public contribution was the release of the first part of the JPEG standard in 1991. Since then, the group has continued to work on both the original JPEG standard and the JPEG 2000 standard (though the latter never became popular).

In the late 1980s, the Moving Picture Experts Group was formed. The purpose was to create a coding standard moving pictures and audio. Since then, the group has developed the MPEG-1, MPEG-2, and MPEG-4 standards.

At the end of the 1990s, a new group called the Joint Video Team (JVT) was formed. It included both the VCEG and MPEG. The purpose was to define a standard for the next generation of video coding. The work was completed in May 2003 and resulted in the MPEG-4 AVC/H.264 standard.

JVT added the new standard to MPEG-4 as a separate part (Part 10) called *Advanced Video Coding*, from which the commonly used abbreviation AVC is derived. The codec was simultaneously launched as a recommendation by the ITU (*ITU-T Recommendation H.264, Advanced video coding for generic audiovisual services*) and as a standard by the ISO/IEC (*ISO/IEC 14496-10 Advanced Video Coding*).

Over the last 10 years, the same organizations that have worked on H.264 have been working both individually and together on a new compression standard. The goal is a 50% reduction in the bitrate for the same subjective image quality as H.264. The new standard is called H.265 or High Efficient Video Coding. In 2013, it was published as an ITU standard, and it later became an MPEG standard.

## 6.3 COMPRESSION FORMATS

---

This section gives a description of the compression formats that are or have been relevant to video surveillance.

First, the JPEG formats are introduced:

- JPEG
- Motion JPEG (MJPEG)
- JPEG 2000

Then, the MPEG formats are introduced:

- H.261 and H.263
- MPEG-1
- MPEG-2
- MPEG-4
- H.264 (MPEG-4 Part 10 AVC)
- H.265

For more information about the technical aspects of the format groups, see Sections 6.4 and 6.5.

### 6.3.1 JPEG

The JPEG standard (ISO/IEC 10918) is the most widely used image compression format today. It is the most common compression format used in smartphones and digital cameras. It is also supported by all web browsers, which makes it widely accepted and easy to use. Users have the flexibility of having high image quality with rather low compression ratio or very high compression



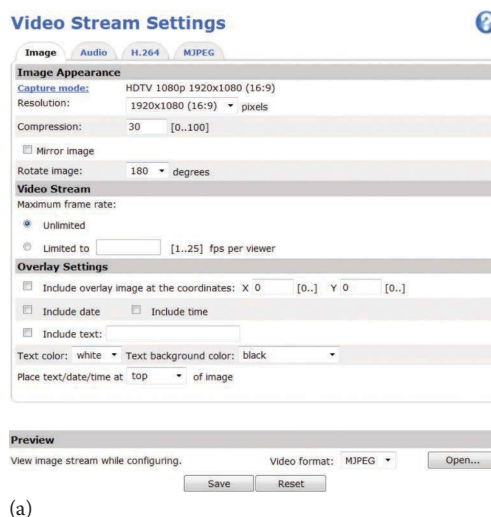
**Figure 6.3** Original image (a) and JPEG compressed image (b) using a high compression ratio that results in blockiness.

ratio with lower image quality. The low complexity of the JPEG technology allows cameras and viewers to be produced at low cost.

Normally, there is no visual difference between a JPEG compressed image and the original uncompressed image. However, if the compression ratio is pushed too high, artifacts in the form of blockiness appear (Figure 6.3).

The JPEG compression standard contains a series of efficient compression techniques. The main technique that actually compresses the image is the discrete cosine transform (DCT) followed by quantization that removes the unnecessary information (the “invisible” parts). The compression level can usually be set as a percentage, from 1% to 99%, where 99% gives the highest compression level and the smallest bitrate but the most artifacts (see Figure 6.4a through e).

For a description of the JPEG compression process, see Section 6.4.



**Figure 6.4** A typical interface for a network camera, in which the compression ratio can be set between 0% and 99%. Depending on the level of detail and complexity in a scene, different compression ratios can be used. In most cases, 50% or more can be used. In some cases, 90% or even 95% can be applied with no or little visible impact (a). *(Continued)*



**Figure 6.4 (Continued)** A JPEG image with no compression (frame size 71 kB) (b). Compare this with the following images: JPEG image with 50% compression (frame size 24 kB) (c). JPEG image with 90% compression (frame size 14 kB) (d). JPEG image with 95% compression (frame size 11 kB) (e).

### 6.3.2 Motion JPEG

MJPEG is a digital video sequence represented as a series of JPEG images. The advantages are the same as with single JPEG still images: flexibility in terms of both quality and compression ratio. In addition, because there is no dependency between the frames, MJPEG is robust. This means that if one frame is dropped during transmission, the rest of the video will be unaffected.

The main disadvantage of MJPEG is that it makes no use of any video compression techniques because it uses only a series of still images. This means a lower compression ratio for MJPEG video sequences than when video compression techniques such as MPEG are used. MJPEG is popular in situations where individual frames in a video sequence are required (e.g., for analysis) and where lower frame rates (typically 2 fps or lower) are used.

### 6.3.3 JPEG 2000

With its better compression ratios, JPEG 2000, or ISO/IEC 15444, was intended to follow the successful JPEG compression technology. The basis was to incorporate advances in image compression research in an international standard. Instead of using the DCT technique, JPEG 2000 uses the wavelet transform algorithm.

The advantage of JPEG 2000 is that the blockiness of JPEG is replaced with an image that is generally “fuzzier,” as seen in Figure 6.5b.



**Figure 6.5** Original image (a) and JPEG 2000 image (b) compressed with high compression ratio to emphasize artifacts.

Whether the fuzziness of JPEG 2000 is preferable over the blockiness of JPEG is a matter of personal preference. For the same image quality, JPEG 2000 enables higher compression ratios than JPEG. For the same image quality at moderate compression ratios, JPEG 2000 produces images typically about 75% of the size of JPEG. However, JPEG 2000 is a far more complex compression technique.

Despite its slight advantage of 25% smaller file size than JPEG, JPEG 2000 never became popular in video surveillance applications, and few web browsers support it.

#### 6.3.4 Motion JPEG 2000

Like JPEG and MJPEG, MJPEG 2000 can also be used to represent a video sequence. The advantage is the same as with JPEG 2000. That is, MJPEG 2000 offers a slightly better compression ratio compared with MJPEG but is more complex.

The disadvantage is similar to that of MJPEG. Because it is a still image compression technique, MJPEG 2000 lacks the advantages of a video compression technique. Therefore, the compression ratios are lower than for video compression techniques.

An undesirable effect with MJPEG 2000 video is that artifacts in the images tend to “float around” between each frame. But with MJPEG the artifacts remain in the same place for each frame in a video stream and look stable over time. Therefore, the viewing experience with MJPEG 2000 is not as good as with an MJPEG stream. MJPEG 2000 has never enjoyed success as a video compression technique.

#### 6.3.5 H.261 and H.263

H.261 and H.263 are not international standards, but recommendations of the ITU. Originally designed for videoconferencing over telephone lines, they are best at low bandwidths and can be seen as simplified versions of MPEG video compression. They are not suitable for use in general digital video coding.

#### 6.3.6 MPEG-1

The first public standard of the MPEG committee was MPEG-1. The first parts were released in 1993. MPEG-1 video compression is based on the same technique used in JPEG and includes techniques for efficient coding of a video sequence. Designed for storage of digital video on CDs, it prioritizes compression ratio over image quality.

### 6.3.7 MPEG-2

The MPEG-2 project focused on extending the MPEG-1 compression technique to handle higher image resolution and quality but that also meant higher bandwidth usage. It provides more advanced techniques for enhancing video quality at the same bitrate but requires more complex and expensive equipment. MPEG-2 is primarily used for storing movies on DVDs.

### 6.3.8 MPEG-4

MPEG-4 (ISO/IEC 14496) is the next generation of MPEG and is based on the same technique as MPEG-1 and MPEG-2.

The most important features of MPEG-4 include the support of low-bandwidth-consuming systems and use cases that require high-quality images and have virtually unlimited bandwidth. The MPEG-4 standard allows for any frame rate, whereas MPEG-2 was locked into 25 fps in PAL and 30 fps in NTSC.

When MPEG-4 is mentioned in video surveillance contexts, it usually means MPEG-4 Part 2. This part is the classic MPEG-4 video streaming standard, also called MPEG-4 Visual.

Some network video streaming systems specify support for *MPEG-4 short header*, which is an H.263 video stream encapsulated with MPEG-4 video stream headers. Because MPEG-4 short header does not take advantage of any of the additional tools specified in the MPEG-4 standard, it produces low-quality video streams.

### 6.3.9 H.264

MPEG-4 Part 10 AVC/H.264, from now on called H.264 in this book, first became available in network video products in 2008. H.264 was jointly defined by standardization organizations in the telecommunications and IT industries and is currently the most widely adopted standard.

The standard addresses several weaknesses in earlier MPEG standards, such as good video quality at substantially lower bitrates, better error robustness, and better video quality at an unchanged bitrate. The standard is designed to give lower latency and better quality for higher latency.

An additional goal of H.264 was to give enough flexibility so that the standard could be used in a wide range of situations and applications with very different demands on bitrate and latency (Table 6.1).

H.264 helped accelerate the adoption of megapixel/HDTV cameras because the highly efficient compression technology can reduce large file sizes and bitrates generated without compromising image quality.

### 6.3.10 H.265

In 2013, the ITU approved the High Efficiency Video Coding standard, which is also referred to as H.265, and a second version was published in early 2015.

The main advantage of H.265 compared to H.264 is the great advances in compression ratio. It improves compression by a factor of 2 for the same video quality. The standard also supports a bigger range of resolutions, from QVGA (320 × 240) to 4320p. However, it only supports progressive

**Table 6.1** Examples of different video services and their bitrate and latency targets

Service/application/industry	Bitrate	Latency
Entertainment video, including broadcast, DVDs, satellite, and cable TV	1–10 Mbit/s	High
Telecom services	<1 Mbit/s	Low
Streaming services	Low	High



or noninterlaced scanning, which codes the entire frame at once. It does not support interlaced video, which codes half of the lines in a frame and then second half an instant later. For more information about image scanning techniques, see Section 4.4.

While the standard has yet to become established in the surveillance industry and new equipment that meet the standard have a higher initial cost, H.265 offers some significant advantages. Higher resolution can be achieved at decreased data rates, lowering minimum bandwidth requirements and reducing storage needs. However, there are trade-offs. While H.265 provides savings in network bandwidth and storage costs, it demands higher-performance network cameras and monitoring stations.

## 6.4 MORE ON JPEG COMPRESSION

The goal of JPEG compression is to deliver the highest possible quality for a given compression ratio. The more important data must be distinguished from less important data in the image. This process is described in the following steps.

1. *Divide the image into macroblocks*: The image is divided into smaller images, in which each macroblock is  $16 \times 16$  pixels. These blocks can be manipulated in real time by fast processors.
2. *Divide image into components*: The image is then divided into components of  $8 \times 8$  pixels and the color information is separated from the texture. It is possible to compress color blocks more without creating visible artifacts. The color spaces Y (brightness) and CB and CR (color) are normally used for this division.
3. *DCT*: It is a form of lossy compression that converts each frame of the image into the transform domain, thus concentrating the image information. Simple structures in the image are given low values, and complex structures receive high values.
4. *Quantization*: The resulting set of values from DCT describes the image content. The process of quantization compresses and eliminates visual data that are almost imperceptible to the human eye. Reducing the number of symbols in the data stream for the image means the stream can be compressed more, and this in turn means the image file size can be reduced. Quantization cannot be reversed and therefore image quality is affected.
5. *Differential pulse code modulation (DPCM) coding of DC component*: DPCM encodes the difference between each current and previous  $8 \times 8$  block. The differences are likely to be small, so encoding the signal requires fewer bits, thus increasing the rate of throughput.
6. *Run-length encoding (RLE), Huffman, and variable-length integer (VLI) encoding*: A combination of techniques such as Huffman encoding, RLE, and VLI encoding is used to remove redundancies and reduce the amount of data needed to store and reproduce the image.
7. *Bitstream generation and byte stuffing*: The bitstream is created and a special step called byte stuffing adds a "00" to each byte with a value of "FF." JPEG uses "FF" as a marker, and a decoding program can misunderstand the "FF" unless it has the "00."

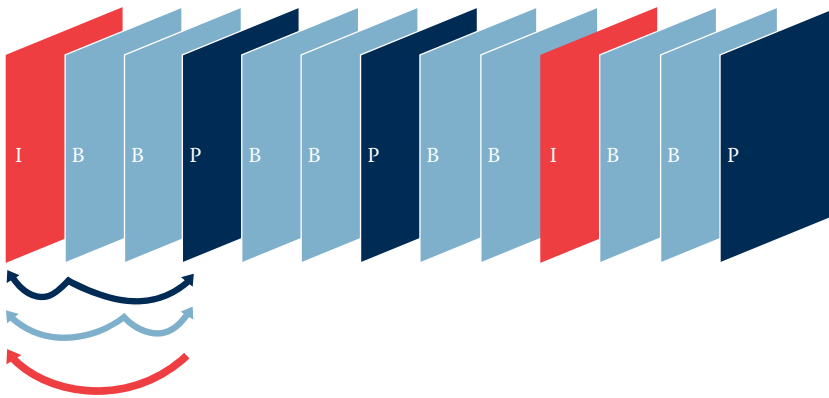
## 6.5 MORE ON MPEG COMPRESSION

The MPEG standard H.264 is the compression standard of choice for most video surveillance applications today, and in the future we can expect to see H.265 emerge as the next popular standard. These fairly complex and comprehensive standards have some characteristics that are vital to understand. The following sections outline these characteristics.

### 6.5.1 Frame types

The basic principle for video compression is image-to-image prediction. The first image in the stream is called an I-frame and is self-contained, having no dependency outside that image. The following frames can use part of the first image as a reference (see Figure 6.6). An image predicted





**Figure 6.6** Typical sequence with I-, B-, and P-frames. A P-frame may only reference preceding I- or P-frames, while a B-frame may reference both preceding and subsequent I- or P-frames.

from one reference image is called a P-frame, and an image that is bidirectionally predicted from two reference images is called a B-frame.

- *I-frames*: Intrapredicted, self-contained.
- *P-frames*: Predicted from last I- or P-reference frame.
- *B-frames*: Bi-directional, predicted from two references—one in the past and one in the future—and thus out-of-order decoding is needed.

The video decoder (“the playback algorithm”) restores the video by decoding the bitstream frame by frame. Decoding must always start with an I-frame, which can be decoded independently, whereas P-frames and B-frames must be decoded together with the current reference image or images.

### 6.5.2 Group of pictures

One parameter that can be adjusted in H.264 is the group of pictures (GOP) length and structure, also referred to as group of video in some MPEG implementations (Figure 6.7). It is normally repeated in a fixed pattern, for example:

GOP = 4 (IPPP IPPP...)

GOP = 15 (IPPPPPPPPPPPPP IPPPPPPPPPPPPPP...)

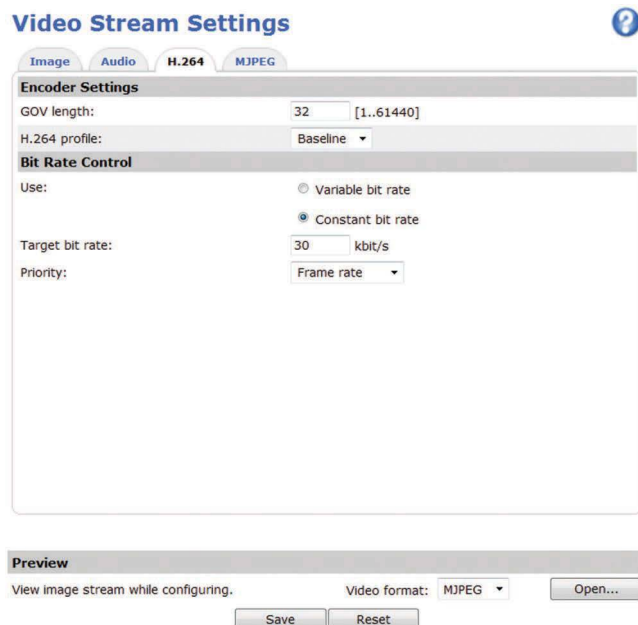
GOP = 8 (IBPBPBPB IBPBPBPB...)

The appropriate GOP depends on the surveillance context. Decreasing the frequency of I-frames decreases the bitrate. Removing the B-frames reduces the latency.

There is a trend toward using longer GOPs of up to 64, 128, and higher because H.264 offers a technique for avoiding error accumulation. This reduces the number of I-frames (which consume a higher bitrate) that are needed and makes the encoding more efficient. The video management software can set limits for GOP length and require 1 fps to maintain access points in the system. Keeping these points available comes at the expense of bitrate.

### 6.5.3 Constant, maximum, and variable bitrates

Another central aspect of MPEG is the ability to choose a bitrate control mode. In most MPEG and H.264 systems, it is possible to choose constant bitrate (CBR), maximum bitrate (MBR), or variable bitrate (VBR). The optimal selection depends on the surveillance situation and available network infrastructure. Also, different camera manufacturers use different technologies and names, but the basic principles are similar.



**Figure 6.7** Interface in a network camera where the length of the group of pictures (GOP), the number of frames between two I-frames, can be adjusted to fit the surveillance context.

When the available bandwidth is limited and stable continuous streaming is essential, the preferred mode is normally CBR because it has CBR or MBR because it has MBR. MBR is sometimes called VBR with a cap.

The goal of CBR is to stay at the target bitrate no matter what happens in the scene. The quality is consistent when the scene complexity is low. But as the activity increases, the camera either drops frames or increases compression. Some CBR techniques also allow padding of data to fill up bitrate gaps in videos of low-complexity scenes. Bit padding is a waste of bandwidth and storage for no gain in video quality.

The goal of MBR is to stay under the target bitrate and to continue to stream video no matter what happens in the scene. Like CBR, MBR uses compression and frame rate to control the stream. It works mainly with compression but drops frames if it has to. MBR reacts quickly to scene changes so when the scene complexity is high, it actively limits bitrate overshoots. When the scene complexity is low, the image quality is high and the bitrate stays below the target. No bit padding is done to reach the target bitrate. When limited bandwidth is available, MBR is typically the best choice for video surveillance.

Some cameras with MBR or CBR have the ability to set a priority between reducing frame rate or quality. Not setting a priority normally means the frame rate and image quality are more or less equally affected.

With VBR, the image quality is consistent regardless of the scene complexity. VBR is suitable in video surveillance situations where images must have a high level of details, especially when there is motion in the scenes. Because the bitrate can vary a lot with VBR, the network infrastructure for such a system needs to have higher bandwidth and storage capacities. In other words, VBR gives crisp details at the price of bandwidth and storage.

The mode and settings that work best depend on the scenario, quality demands, and network infrastructure:

- *Is the scene complexity high or low?* High-complexity scenes require more bandwidth. Medium or high motion means high complexity. Little or no motion means low complexity. Because night-time videos tend to have more noise, they usually need more bandwidth than daytime video.

- *Is frame rate or image quality more important?* A very compressed image has fewer details, making identification more difficult. A very low frame rate means that some activity may go undetected because too many critical frames drop.
- *What is the bandwidth and storage availability?* If ample bandwidth and storage is available, and details in the video are essential, use VBR or set the target bitrate very high. Limit the bandwidth too much and the image quality suffers. Limit the bandwidth to little and waste money on storage.

#### 6.5.4 Profile@Level

Because both MPEG-2 and MPEG-4 cover a wide range of image sizes, frame rates, and variable bandwidth usage, MPEG-2 introduced a concept called Profile@Level. It was created to make it possible to only support subsets of the standard and to communicate compatibilities among systems. Examples of common profiles are MPEG-2 Main profile at Main Level (MP@ML), MPEG-4 Main profile at L3 Level, and H.264 Main profile at Level 5.

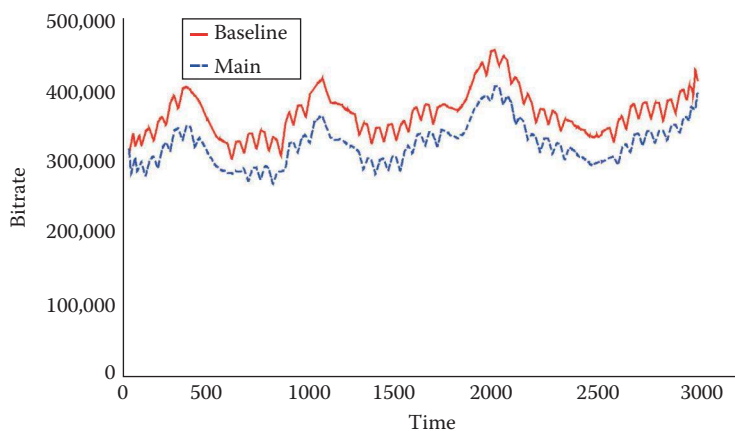
The Internet Streaming Media Alliance uses the Profile@Level definitions to ensure that devices streaming video over the internet are compatible.

#### 6.5.5 Baseline and main profiles

One aspect of H.264 to consider is its various profiles. The two most popular profiles in H.264 encoding for video surveillance applications are the Baseline and Main profiles. The Baseline profile for H.264 uses only I- and P-frames, while the Main profile may also use B-frames in addition to I- and P-frames. The Baseline profile allows network video products to have low latency. In video products with more powerful processors, the Main profile is sometimes used without B-frames to enable higher compression and at the same time low latency and maintained video quality. Using Main profile H.264 compression, VGA-sized video streams can be reduced by 10%–15% and HDTV-sized video streams can be reduced by 15%–20%, compared to using Baseline profile H.264 compression (see Figure 6.8).

The most significant part of the Main profile is the new entropy coder called context-adaptive binary arithmetic coding (CABAC), which replaces the simpler context-adaptive variable-length coding (CAVLC) used in the Baseline profile. CAVLC and CABAC produce the same video quality, but CABAC can do it with reduced frame rate without introducing new artifacts.

The standardization organization continues to add new profiles to H.264. However, not all profiles are suitable for video surveillance.



**Figure 6.8** A comparison of bitrates generated by Baseline profile and Main profile H.264 compression. Main profile can offer higher compression with low latency, while maintaining image quality.

### 6.5.6 Improving H.264 for surveillance needs

Because the H.264 standard dictates the method of decoding of video rather than the encoding, developers have the freedom to improve the encoding solutions as long as they maintain playback compatibility (see also Section 6.6). An example of such an improvement is a collection of algorithms that work together to analyze the video stream in real time. Areas with interesting details and motion are preserved at the given video quality, while low-interest areas can be filtered and compressed more aggressively.

One of the algorithms, dynamic GOP, removes redundant I-frame updates. By switching between a maximum GOP value for busy scenes and a default value for low-activity scenes, the bitrate of the video can be drastically reduced.

Another algorithm optimizes bitrate in real time by analyzing where the available bits will have the most benefit from a forensic standpoint. The process repeats region by region and frame by frame. Depending on what happens in the scene, this dynamic region of interest (ROI) automatically changes shape and size, appears and disappears, and splits and merges. Therefore, dynamic ROI is better prepared for unexpected events than a traditional ROI implementation where the region is set manually.

Again, these improvements fully comply with the standard because they only change how the video is encoded but not its playback. Some improved implementations of H.264 could yield bitrate savings of 50% or more and therefore become a more viable option than early implementations of H.265. The benefit with improved H.264 is that because it is just a firmware upgrade away, existing network cameras can be used, and the same video management system and monitoring station can be used without changes. Figure 6.9 shows an example of an implemented improvement of H.264.

### 6.5.7 Licensing

MPEG-2, MPEG-4, and H.264 are subject to licensing fees, which any company manufacturing products using these compression standards must pay. MPEG LA, an independent license administrator, manages the licensing fees. For most network video products, one or several licenses have been paid by the manufacturer, which means that the video can be viewed on one or a few monitoring stations.

An end user planning to view the video at more monitoring stations than the product is licensed for must buy additional licenses to match the number of stations. If the manufacturer has not paid



**Figure 6.9** The same scene captured with H.264 (left) and improved H.264 (right). In this case, improved H.264 results in bitrate savings of nearly 60% with similar quality video from a video surveillance perspective.

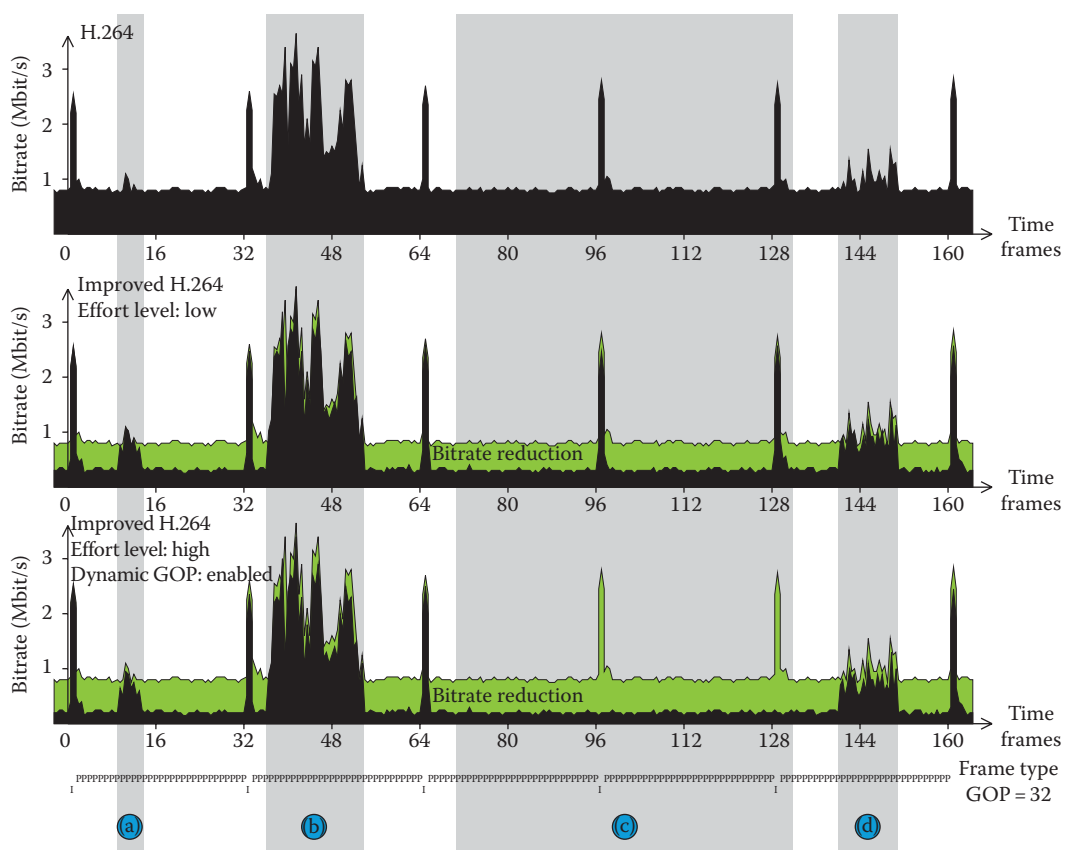
the license fees, it most likely means that the manufacturer does not fully follow the compression standard, which in turn limits the compatibility with other systems.

### 6.5.8 Backward compatibility

MPEG-2 and later standards are not backward compatible. This means that decoders and encoders that strictly comply with MPEG-2 do not work with MPEG-1. Likewise, H.264 encoders and decoders do not work with MPEG-2 or previous versions of MPEG-4 unless they are specifically designed to handle multiple formats. However, various solutions are available where streams encoded with newer standards can sometimes be packetized inside older standardization formats to work with older systems.

## 6.6 COMPARING STANDARDS

When comparing the performance of MPEG standards such as MPEG-4 and H.264, it is important to note that results may vary between encoders that use the same standard. This is because the designers of encoders can choose to implement different subsets of the standard. As long as the output of an encoder conforms to a standard's format and decoder, it is possible to make different implementations. This helps to optimize the technology and reduce the complexity in



**Figure 6.10** This example highlights the behavior of improved H.264 under different conditions: A period with small short-lasting movement. The small motion is detected, and adding bits in that region can preserve the moving part of the video (a). A period with large long-lasting movement needs more space, but still it is possible to save storage during this motion period because the dynamic region detects areas where non-prioritized information can be removed (b). Periods without motion are detected, and the dynamic GOP algorithm removes unnecessary I-frame updates (c). Periods with small long-lasting movement (d).

implementations. However, it also means that an MPEG standard cannot guarantee a given bitrate or quality and comparisons cannot be properly made without first defining how the standards are implemented in an encoder.

Unlike an encoder, a decoder must implement all the required parts of a standard to decode a compliant bitstream. This means that only the decoder is truly standardized. The standard specifies exactly how a decompression algorithm should restore every bit of a compressed video. If video quality is a concern, the end user should buy and test a few products to make sure that the quality matches the purpose.

Given the same level of image quality, Figure 6.10 shows a bitrate comparison between H.264 and improved H.264.

For more information about the Baseline and Main profiles, see Section 6.5.6.

## 6.7 BEST PRACTICES

One compression standard and configuration does not fit all situations. When designing a network video application, consider the following questions:

- *Frame rate*: What is the required frame rate? Is the same frame rate required all the time? Below 2 fps, consider using MJPEG and controlling the frame rate through motion detection. For higher frame rates, H.264 is normally best because it saves bandwidth and storage.
- *Bandwidth*: What is the available network bandwidth? In scenarios with very low bandwidth, H.264 compression using CBR or MBR bitrate may be the only option, but image quality will be sacrificed when motion occurs in the scene.
- *Image quality*: What is the allowed level of image degradation (artifacts) due to compression? Compression ratios well above 90% can be used if the scene is not too complex.
- *Latency*: What is the acceptable level of latency? If video is not monitored live but only recorded, latency might not be an issue. When controlling PTZ cameras, it is important to have low latency.
- *Bitrate control*: How much bandwidth is available? Can VBR be used, or is CBR or MBR the best choice? If CBR or MBR is used, is it possible to set the priority for whether frame rate or image quality should be reduced?
- *Robustness*: How reliable or secure must the system be? Is it acceptable that video might be lost for 0.5 seconds if a frame is dropped on the network?
- *Standard and compatibility*: Is the openness of the system and interoperability with other systems essential? If so, make sure the chosen products follow the standard 100%.



## CHAPTER 7

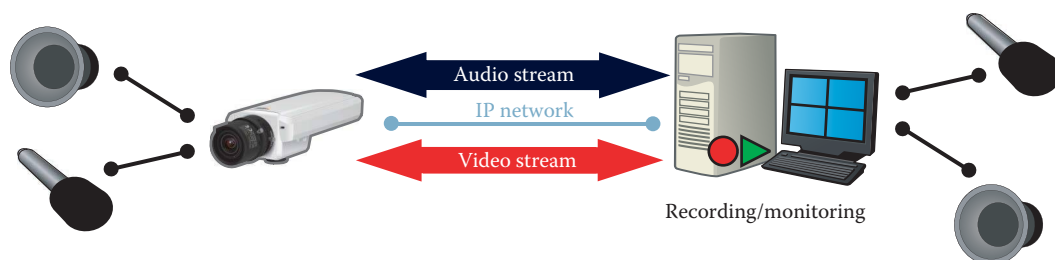
### Audio technologies

Much of our learning and our relations with others are conducted through audiovisual cues—that is, through what we see, hear, and say. In our daily lives, we often are alerted to unusual events first by what we hear. Then we verify the events visually.

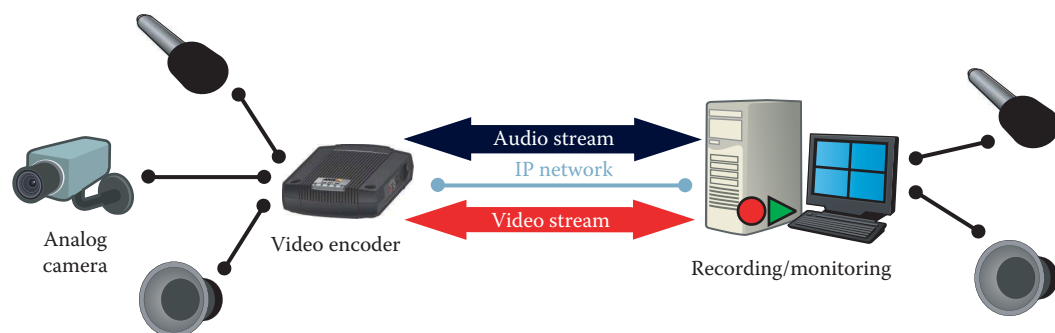
Integrating audio with a video surveillance system can be invaluable to the system's capabilities to, for example, detect and interpret events and emergencies. Consider a video surveillance system without audio. If the source were outside the camera's field of view, a cry for help, the sound of breaking glass, a gunshot, or an explosion in the vicinity would escape notice by a video surveillance system without audio. Even if a scene such as a building entrance were under visual surveillance, without audio support the system would not be able to pick up the sound of breaking glass or any sound at all. The ability of audio to cover a 360° area enables the video surveillance system to extend its coverage beyond a camera's field of view. It can instruct a pan, tilt, and zoom (PTZ) camera or alert the operator of one, to verify an audio alarm visually. Because it gives more information to remote users and speaks to another sensory system, an audiovisual surveillance system increases the effectiveness of the security solution.

Audio can be used in other ways too. It can give users the capability of both listening in on an area and communicating with visitors or giving orders and warnings to intruders. For example, if a person in a camera's field of view demonstrates suspicious behavior, such as loitering near a bank machine, or is seen entering a restricted area, a remote security guard can send a verbal warning to the person. In a situation where a person has been injured, the person can be greatly comforted by a voice that can keep them alert and assure them that help is on the way. Audio can also be used in access control applications—that is, a remote “doorman” can communicate with visitors at an entrance. Other applications include videoconferencing and remote helpdesk scenarios, such as a parking garage without attendants.

Although having audio in a video surveillance system is still not widespread, its implementation is expected to increase along with the growing adoption of network video. Network video enables easier implementation of audio than analog closed-circuit television (CCTV). In an analog system, separate audio and video cables must be installed from endpoint to endpoint, meaning that the cables run from the camera and microphone location to the viewing or recording location. If the distance between the microphone and the station is too long, balanced audio equipment must be used, which increases installation costs and difficulty. In a network video system (Figure 7.1), a network camera with audio support processes the audio and sends both the audio and video over the same network cable for monitoring and recording. This eliminates the need for extra cabling and makes audio and video synchronizing much easier.



**Figure 7.1** A network video system with integrated audio support. Audio and video streams are sent over the same network cable.



**Figure 7.2** Some video encoders have built-in audio, which makes it possible to add audio even when analog cameras are used in an installation.

Support for audio can be found in many types of network cameras, including fixed, PTZ, and fixed dome cameras. Many camera manufacturers are recognizing the importance of audio, and audio is becoming a common feature in network cameras.

Some video encoders also have built-in support for audio, which means that they can provide audio functionality in an analog camera installation. This may be useful in an application where a specialty camera is used or if existing analog cameras are installed. Because the video encoder can be located close to the analog camera, the length of the audio cables can be shorter (Figure 7.2). Another audio device for network video applications is the audio module, which only has audio support. The audio module also has input/output (I/O) ports that can be located far away from a network camera. For example, in a city surveillance application, the audio module and microphone are located close to street level, while the PTZ camera is located high up on a pole.

When considering to implement audio, its application should be clear because it affects what products should be selected. The following sections of this chapter discuss audio transmission modes (simplex, half duplex, or full duplex), audio equipment (microphones, speakers, and cabling), acoustical adjustments, audio detection alarms, codecs and bitrates, audio and video synchronization, and a summary of the factors that affect audio quality. For information about how the use of audio can sometimes be restricted or regulated by local legislation or codes of practice, see Chapter 17.

## 7.1 AUDIO MODES

Depending on the application, audio might need to be transmitted in one direction or both directions. This can be done either simultaneously or in one direction at a time. There are three basic modes of audio communication: simplex, half duplex, and full duplex.

### 7.1.1 Simplex

Simplex means that audio can be sent in one direction only. Audio is sent either from the camera—which is most often the case—or from the user. Situations where audio is sent only from the camera include remote monitoring and video surveillance applications where live audio, as well as video, from a monitored site is sent over a network (Figure 7.3). Applications where audio is sent only from a user or operator include situations where there is a need to speak instructions to a person detected by the camera or in parking lot scenarios where the operator can use audio to scare off a potential car thief (Figure 7.4).

### 7.1.2 Half duplex

Half duplex means that audio can be sent and received in both directions—from the camera and the operator—but only in one direction at a time (Figure 7.5). This type of communication is similar to a walkie-talkie. To speak, an operator must press and hold down a push-to-talk button. Releasing the button enables the operator to receive audio from the camera. With half duplex, there is no risk of echo problems (a topic discussed in Section 7.3.3).

### 7.1.3 Full duplex

Full duplex means that users can send and receive audio (talk and listen) at the same time (Figure 7.6). This mode of communication is similar to a telephone conversation. Full duplex requires the client PC to be able to handle full-duplex audio. While full duplex has the advantage of simultaneous audio in both directions, it also increases the demands on available bandwidth.



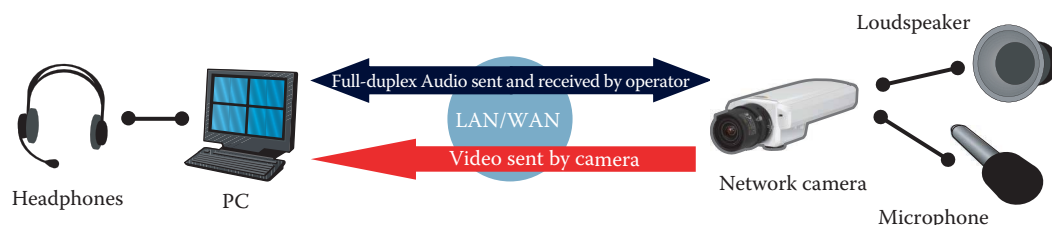
**Figure 7.3** In simplex mode, audio is sent in one direction only. In this case, the camera sends audio to the operator.



**Figure 7.4** In simplex mode, audio is sent in one direction only. In this case, the operator sends audio to the camera.



**Figure 7.5** In half-duplex mode, audio is sent in both directions, but only one party at a time can send.



**Figure 7.6** In full-duplex mode, audio is sent to and from the operator simultaneously.

## 7.2 AUDIO EQUIPMENT

When a network camera or a video encoder has support for audio, it may very well include a built-in microphone but rarely a built-in speaker. The built-in microphone may be appropriate in some surveillance contexts, but in many cases an external microphone is a better solution. This section gives some guidance when selecting audio equipment.

### 7.2.1 Audio input (microphones)

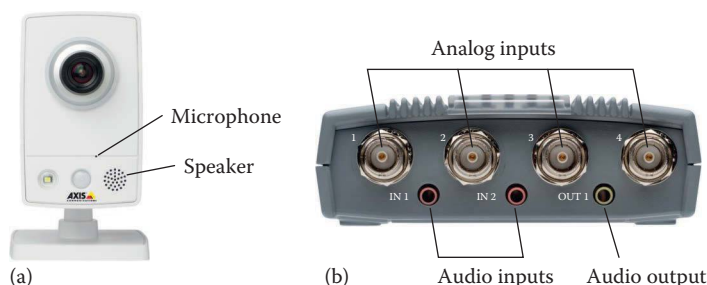
A network camera or video encoder with integrated audio functionality often has a built-in microphone or a mic-in/line-in jack (Figure 7.7). A camera with mic-in/line-in support gives users the option of using another type or quality of microphone than the one built into the camera. In addition, a microphone can be located some distance away from the camera. For example, the camera can be located close to the ceiling, while the microphone can be placed close to a door.

Audio sources can have different output signal levels. For example, there are microphones with and without built-in amplifiers. In many cases, the camera supports the use of both, allowing connection of a microphone without an amplifier to mic-in and a microphone with an amplifier to line-in. Often the same jack is used, and the choice between mic-in and line-in is made in the software. Line-in means the network camera can be connected to devices that deliver an already amplified audio signal (known as line signal). Devices that provide line signals include mixers, which allow a network camera to connect to several microphones, and microphone amplifiers, which connect to microphones without built-in amplifiers.

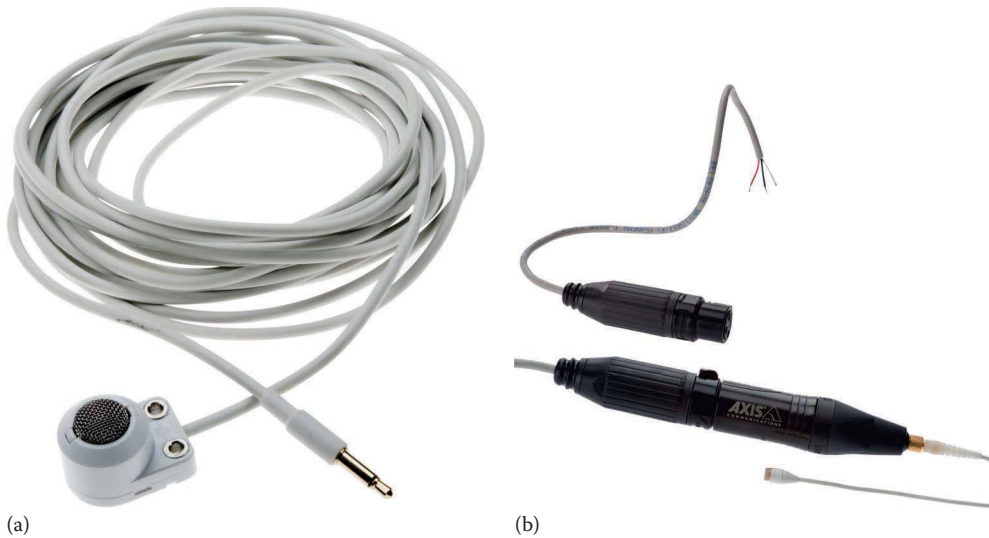
There are three main types of microphones: condenser, electret condenser, and dynamic. They differ in the way they convert sound into electrical signals. See Sections 7.2.1.1 through 7.2.1.3.

#### 7.2.1.1 Condenser microphones

Of the three main types of microphones, the condenser has traditionally been considered to offer the highest audio sensitivity and quality, though some back-electret microphones can deliver similar quality. It is often used in professional recording studios and can be used in video surveillance applications that require high audio quality. A condenser microphone is used together with



**Figure 7.7** A camera (a) with built-in microphone and speaker and a video encoder (b) with inputs for external microphone and speakers.



**Figure 7.8** A condenser microphone with a 3.5 mm connector (a). A condenser microphone, a phantom power input, and a pigtail adapter (b). The condenser microphone (bottom right) connects to the phantom power input, which has a male XLR connector in its other end. On the top right, the pigtail adapter's female XLR connector is visible.

a so-called phantom power supply, which supplies the necessary power required by the microphone. Sometimes a condenser microphone uses an XLR connector, a circular connector with three pins often found on professional audio equipment. If a network camera does not support an XLR connector, a condenser microphone can still be connected to the camera through an adapter (Figure 7.8b), which sometimes is provided by a phantom power supply box.

#### 7.2.1.2 Electret condenser microphones

The built-in microphone in a network camera is often an electret condenser microphone. This type of microphone is common in headsets and computer microphones (Figure 7.9) and often uses a 3.5 mm audio connector. It offers a high level of sensitivity and is less expensive than a condenser microphone. The electret condenser microphone normally needs a voltage of 1–10 V. If an external electret microphone is used with a network camera, the network camera can supply the microphone with the necessary power.



**Figure 7.9** A computer headset with a microphone based on electret condenser technology.

### 7.2.1.3 Dynamic microphones

The dynamic microphone is rarely used in the video surveillance industry because its audio sensitivity is not high enough. It also has a poor ability to reproduce low frequencies when the source of the sound is not close to the microphone. The dynamic microphone often uses an XLR connector. If a network camera does not support an XLR jack, an adapter can be used to connect the camera with a dynamic microphone.

### 7.2.1.4 Directional microphones

Microphones are made with different polar patterns, also called pick-up or directional patterns. The pattern types include omnidirectional, which picks up audio equally in all directions, and unidirectional such as cardioid (which means heart shaped) and supercardioid, which has high audio sensitivity in one specific direction. To pick up sounds in a specific spot located far away from a network camera, a specialized unidirectional microphone called a shotgun microphone can be used.

## 7.2.2 Audio output (speakers)

There is a wide array of speakers available that can be used along with network cameras. PC speakers are used most often. The power of speakers is measured in watts. It indicates how much power the speakers consume and often relates to how loud the speakers can be, as well as the speakers' sensitivity specification.

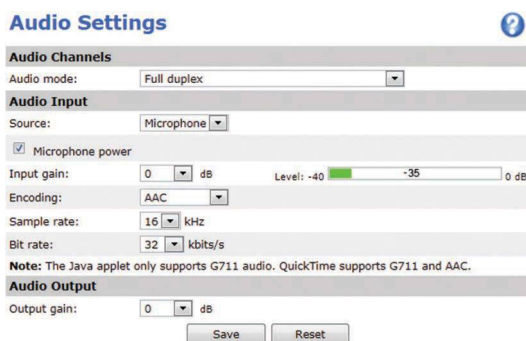
An active speaker, which is a speaker with a built-in amplifier, can be connected directly to a network camera. If a speaker has no built-in amplifier, it must first connect to an external amplifier that is connected to a network camera.

## 7.3 ACOUSTICAL ADJUSTMENTS

There are a number of adjustments (Figure 7.10) that can be made to get the best audio performance from an installation. For example, it can include methods for reducing the volume of unwanted sounds and amplifying quiet sounds. This section discusses several of the most common adjustments, such as volume and gain, audio processing, echo cancellation, and noise reduction.

### 7.3.1 Volume and gain

The level of audio delivered from a microphone can be adjusted by tuning the gain on the amplifier, which can be built in or connected to the microphone. To optimize the audio quality, the gain setting on the camera (if using a camera's built-in microphone) or on the amplifier (if using a stand-alone microphone) should be adjusted so that the signal is not clipped or distorted. It is crucial that the gain level is chosen wisely and that necessary gain is applied in the signal path as early as possible, preferably in the microphone if available.



**Figure 7.10** An example of a user interface for audio settings on a network camera.





**Figure 7.11** A user interface showing settings for audio processing, echo cancellation, and noise reduction.

### 7.3.2 Audio processing

Audio processing is a group term for a number of different types of automatic processes that work together to correct and control audio signals. For example, it can include equalization and automatic gain control. Equalization adjusts the balance between different frequencies and automatic gain control amplifies weak signals. When all the audio processing is done, echoes have been canceled, and noise reduced, the audio signal can be sent back without reverberating sounds or other distortions.

### 7.3.3 Echo cancelation

In full-duplex mode, the microphone captures not only the desired incoming sounds but also the sounds produced by a built-in loudspeaker. Such sounds, known as feedback, can be reduced by facing the microphone away from the loudspeaker and using echo cancellation (Figure 7.11). A network camera with echo cancellation has a signal processor with a short-time memory that remembers the audio signals that have just been sent from a loudspeaker or that leak between neighboring wires. If the microphone picks up such audio signals, the device recognizes the signals as an echo and removes them.

### 7.3.4 Noise reduction

Noise reduction (Figure 7.11) can be used to reduce background noise. This feature is configured using two parameters: threshold and attenuation. Threshold is used to define the level under which noise will be reduced. Then attenuation can be used to choose the degree of noise reduction. Full background noise reduction may not be desirable because a listener might interpret it as a break in the connection. Because noise reduction can decrease the quality of the audio, it should be used carefully. There are also other methods to reduce overall noise in audio during recording of the desired signals.

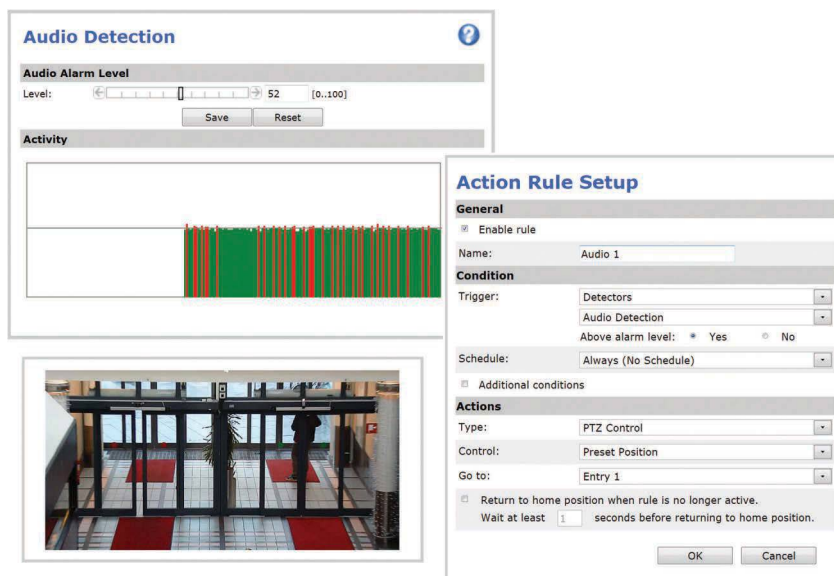
## 7.4 AUDIO DETECTION ALARM

Just like network camera can analyze video, it can analyze audio. In areas that are too dark to effectively detect motion in video, audio detection can be used as a complement to motion detection. It can also be used to detect activity in areas outside the camera's view.

When sounds are detected, such as the breaking of a window or voices in a room, they can trigger a network camera to send and record video and audio, send email or other alerts, and activate external devices such as alarms. Similarly, alarm inputs such as motion detection and door contacts can be used to trigger video and audio recordings.

In a PTZ camera, audio detection alarm can trigger the camera to turn automatically to a preset location such as a specific door or window (see Figure 7.12).

Depending on need, audio detection can be used all the time or during specific times, or it can be turned off altogether. It can be set to trigger an event if the incoming sound level rises above, falls below, or passes a certain level of sound intensity.



**Figure 7.12** When the audio input reaches the audio alarm level, it triggers the PTZ camera to turn to a preset position or to do a guard tour. It helps an operator both to detect sound and to verify whether the sound is cause for further investigation.

## 7.5 AUDIO COMPRESSION

Analog audio signals must be converted into digital audio through a sampling process. Then it is almost always compressed to reduce the size for efficient transmission and storage. The sampling to digitize the audio is often done in a hardware codec, while the actual compression is done using an audio codec, a software algorithm that codes and decodes audio data. The following sections present some factors that can influence audio quality and file size.

### 7.5.1 Sampling rates

Sampling rate or frequency refers to the number of times per second a sample of an analog audio signal is taken and is defined in terms of hertz (Hz). The human ear can hear sounds up to 20 kHz, but to capture this level of sound with good quality, a sampling rate of at least 40 kHz is necessary. Music CDs, for example, use a sample rate of 44.1 kHz.

The sample rate (according to the Nyquist–Shannon sampling theorem) must be at least twice the maximum required frequency. If the human voice is the only sound of interest, a sampling rate of at least 8 kHz is needed because the frequency of the human voice is normally below 4 kHz. In general, the higher the sampling rate, the better the audio quality, and the higher the demands on bandwidth and storage.

### 7.5.2 Bitrate

The bitrate setting is important because it determines the level of compression and thereby the quality of the audio. In general, the higher the compression level, the lower the audio quality. Regarding audio quality, the differences between software codecs can be particularly evident at high compression levels (low bitrates) but not at low compression levels (high bitrates). Higher compression levels can also introduce more latency or delay, but again that needs to be weighed against the savings on bandwidth and storage.

The most common bitrates for various codecs range between 32 and 64 kbit/s. Newer codecs produce adequate sound quality for bitrates as low as 32 kbit/s. Audio bitrates, like video bitrates, are important to consider when calculating total bandwidth and storage requirements.

Some codecs feature only a constant bitrate (CBR) mode, whereas others enable both CBR and variable bitrate (VBR) modes. When using VBR, the bitrate adjusts to the complexity of the audio. This means that less demanding audio is compressed more and generates lower bitrates than more complex sounds do. This enables delivery of a higher-quality stream than a CBR file of the same size. When using VBR, a target bitrate can be set so that the levels of compression fluctuate close to the desired bitrate. The downside of using VBR is that the encoding time may be longer.

### 7.5.3 Software audio codecs

Today there are three main audio codecs used in network video:

1. *AAC-LC (Advanced Audio Coding—Low Complexity)*: Requires a license and supports either constant or VBR.
2. *G.711*: Is a nonlicensed technology and only supports CBR.
3. *G.726*: Is a nonlicensed technology and only supports CBR.

A fourth codec that is applicable for network video is G.722.2. More details on each of the four codecs follow.

#### 7.5.3.1 AAC-LC

The official name of AAC-LC is MPEG-4 AAC. It includes four different profiles, where LC profile is the least complicated and the most widely used form. AAC-LC requires a license for encoding and decoding.

AAC offers sampling rates ranging from 8 to 96 kHz and bitrates ranging from as low as 2 kbit/s for low bitrate speech encoding to more than 300 kbit/s for high-quality audio coding. It supports constant and VBR modes.

If achieving the best possible audio quality is a priority, AAC is the recommended codec to use, particularly at a sampling rate of 16 kHz or higher and at a bitrate of 64 kbit/s. If a product does not offer a sample rate of 16 kHz or higher, then AAC with a sample rate of 8 kHz at a bitrate of 24 or 32 kbit/s is recommended.

AAC-LC was developed by a group of companies, which includes Dolby, Fraunhofer IIS, Sony, and AT&T, and has been part of the MPEG standard since 1997. It is specified as Part 7 of the MPEG-2 standard and Part 3 of the MPEG-4 standard. Because AAC-LC is part of the MPEG-4 standard, it is very possible that it will become the most adopted standard in the video surveillance industry. For more information on the MPEG group, see Chapter 6 about compression technologies.

#### 7.5.3.2 G.711 PCM

G.711 PCM (pulse-code modulation) is an unlicensed speech codec from the ITU's Telecommunication Standardization Sector (ITU-T). It has lower delay and requires less computing power than AAC-LC. It was developed in 1972 as a telephony standard. All IP telephony and voice over IP (VoIP) manufacturers support this standard. Therefore, it is very useful when integrating audio into a VoIP system. G.711 PCM has a sampling rate of 8 kHz and a bitrate of 64 kbit/s. It has a rather poor quality-to-bitrate ratio, but it is still used in some systems.

With G.711, it is important that the client also use  $\mu$ -law compression, which is a technique that takes a 14-bit signed linear audio, increases the magnitude by 32, and converts it to an 8-bit value.

#### 7.5.3.3 G.726 ADPCM

G.726 ADPCM (adaptive differential pulse-code modulation) is an unlicensed speech codec from the ITU-T. G.726 is a low-power and low-cost implementation standard with low latency. Like G.711, G.726 has lower delay and requires less computing power than AAC-LC. It has a sampling rate of 8 kHz and bitrates of 16, 24, 32, and 40 kbit/s. The most commonly used bitrate is 32 kbit/s.

It is the most commonly used codec within the security industry, but it is not widely used elsewhere. G.726 ADPCM was introduced in 1990.

#### 7.5.3.4 G.722.2 or AMR-WB

G.722.2 or AMR-WB (adaptive multirate—wideband) is a licensed speech codec from the ITU-T. It has a sampling rate of 16 kHz and offers bitrates ranging from 7.60 to 23.85 kbit/s. This is the codec used in networks such as UMTS, a 3G mobile phone technology. G.722.2 offers good speech performance at rates of 12.65 kbit/s and higher.

## 7.6 AUDIO AND VIDEO SYNCHRONIZATION

---

A media player (a computer software program used for playing back multimedia files) or a multimedia framework (such as Microsoft® DirectX®, a collection of application programming interfaces that handle multimedia files) manages the synchronization of audio and video data. Audio and video are two separate packet streams that are sent over a network. For the client or player to synchronize the audio and video streams perfectly, the audio and video packets must be time-stamped.

Some network cameras may not support time stamping of video packets when using Motion JPEG compression. If this is the case and if having synchronized audio and video is critical, it is better to use MPEG-4 or H.264 compression because such video streams are usually sent using Real-Time Transport Protocol (RTP), which time-stamps the video packets. In the future, H.265 will also become an alternative in this scenario.

However, there are many situations where synchronized audio is less important, such as when audio is monitored but not recorded.

## 7.7 THE FUTURE OF AUDIO IN NETWORK VIDEO

---

Audio is likely to become a key concern when selecting a video surveillance solution in the future. When used together, audio and video offer a more complete monitoring solution.

Future network cameras will likely support dual audio codecs simultaneously, just as two video compression formats (H.264 and Motion JPEG) can be supported simultaneously. This will enable users to take advantage of the different strengths of the codecs and apply them for different purposes, for example, one codec for recording and another for communications.

Other improved functionalities will likely lie in the following areas: synchronization between audio and video, audio analytics (real-time or post-event), audio tracking with PTZ cameras, audio quality, and the ability to identify certain sounds such as speech and breaking glass.

## 7.8 OTHER AUDIO DEVICES IN NETWORK VIDEO SYSTEMS

---

As audio continues to be a more important complement to network video and physical security systems, new IP-based audio solutions have lately become available that can augment the systems. Some of those products are network speakers and door stations, which are often based on the Session Initiation Protocol (SIP).

### 7.8.1 Network speakers

The difference between a network speaker (see Figure 7.13) and a regular speaker is the same as between an analog camera and a network camera. The network speaker connects directly to the network, is powered through Power over Ethernet (PoE), and includes local intelligence. The benefits are ease of install, ease of integration, as well as built-in intelligences such as automatic monitoring of functionality and ability to play prerecorded messages and communicate with a phone or cell phone over the SIP (see Section 7.8.3).



**Figure 7.13** A network horn speaker.

## 7.8.2 Network door station

A network door station (see Figure 7.14) is an IP-based camera with intercom functionality. It is used for two-way communication, video identification, and remote entry control. In simple words, a network door station is a virtual door attendant. People on the inside can see and talk to people on the outside while they remain at their desk or are on the go. Then, through the same interface, they can choose to take a number of different actions such as unlock the door, start a recording, or even lock down the whole building. It is a perfect complement to any surveillance system as it offers effective control of entry to premises and is as easy to install as any network camera.

See Figures 7.15 and 7.16 for an example of a surveillance system with network door stations and a scenario where door stations help retail personnel work more effectively.

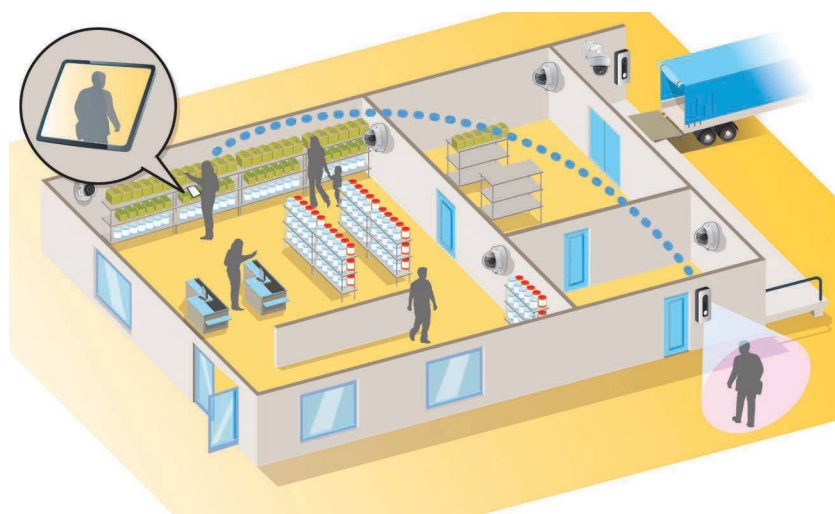
To make calls, a network door station uses VoIP. VoIP is a group of technologies that enables voice communication and multimedia sessions over IP networks. A number of different proprietary and open-source protocols can be used to implement VoIP. One of the most popular open-source protocols is SIP. For more information about SIP, see Section 7.8.3.



**Figure 7.14** A network door station.



**Figure 7.15** Door stations that are connected to a network can communicate with monitoring stations, IP phones, and smartphones or tablets.



**Figure 7.16** Through an IP phone with an LCD screen, a smartphone, or a tablet, the user can identify visitors, such as sales representatives and delivery drivers, and let them in without having to leave their desk, customer, or monitoring station.

The use of IP standards and the open interface makes it easy to integrate a network door station in smaller installations as well as more advanced enterprise systems. It connects to the existing IP network, and thanks to PoE, a single network cable is all it takes to power both the door station itself and a standard door lock. If the door station has I/O and relay connectors, it can be connected to other devices as well, for example, a request to exit device or a safety relay.

### 7.8.3 SIP

SIP is a text-based protocol, similar to HTTP and SMTP, for communication over IP networks. It is used to start, change, and end media stream sessions, which can include voice and video elements. It has become a widely adopted standard protocol for IP telephony, video conferencing, call control, and instant messaging.



SIP calls can be set up in many ways, but there are three main types:

1. Peer-to-peer calls (also called local calls)
2. SIP server calls (also called private branch exchange [PBX] calls)
3. SIP trunk calls

Peer-to-peer calls are calls between two devices (such as computers, softphones, door stations, cameras, IP desk phones) that belong to the same network. The call is made to the SIP address of the device.

To make SIP server calls, the devices must be connected to a SIP server that handles the call exchanges. A SIP server, or a PBX, is a hub that works like a traditional switchboard. It can be hosted on an intranet or by a third-party service provider. The SIP-enabled devices register with the SIP server and can contact each other through their SIP addresses. A PBX can show call status, allow call transfers, handle voicemail, and redirect calls among other things.

SIP addresses (also known as SIP uniform resource identifiers [URIs] or SIP numbers) are used to identify users within a network just as you would use a phone number or an email address to contact a friend or colleague. Like email addresses, SIP addresses are a type of URI that includes two user-specific parts, a user ID or extension, and a domain name or IP address. Together with a prefix and the @ symbol, they make up a unique address. For example, if Caesar of ancient Rome had both an email address and a SIP address, they could be `mailto:caesar@ancientrome.it` and `sip:caesar@ancientrome.it`, respectively. In the case of a peer-to-peer call, the SIP address would include the IP address rather than the domain name `ancientrome.it`.

With a service provider that offers SIP trunking, the traditional telephone network can be used to make calls and traditional phone numbers can be assigned to the SIP devices. This way calls can be made from a network speaker or a network door station to a cell phone or the other way around. Often, providers charge extra for this service.

## 7.9 BEST PRACTICES

---

There are some things to consider when pursuing ideal audio performance in an installation:

- *Audio equipment and placement:* Choose and set up audio equipment based on context, needs, and environment. The microphone type and placement, polar pattern, and cabling and speakers affect audio quality. Although an audio signal can be amplified later, appropriate selection and placement of audio equipment help reduce noise. The microphone should be placed as close as possible to the source of the sound. In full-duplex mode, a microphone should face away and be placed some distance from speakers to avoid feedback.
- *Amplify the signal as early as possible:* This minimizes noise in the signal chain. In addition, make sure the signal levels are as close to the clipping level but not over it. The clipping level is the signal level at which audio is distorted.
- *Apply appropriate signal processing technologies to improve audio quality:* Audio quality can be improved by adjusting the input gain and using different features such as echo cancellation and speech filter.
- *Select the right codec and bitrate:* The codec and bitrate affect audio quality. In general, the higher the bitrate, the better the audio quality.
- *Use shielded cables:* To minimize disturbance and noise, always use a shielded audio cable and avoid running the cable near power cables and cables carrying high-frequency switching signals. Also, keep audio cables as short as possible. If a long audio cable is required, balanced audio equipment (the cable, amplifier, and microphone) should be used to reduce noise.
- *Understand legal implications:* Some countries restrict the use of audio and video surveillance. It is a good idea to check with the local authorities before investing in a system.



# Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

## CHAPTER 8

### Video encoders

Video encoders, also sometimes called video servers, are key pieces of equipment that help analog closed-circuit television (CCTV) systems migrate into an open platform-based network video system. They will continue to play a significant role in the video surveillance market because the vast majority of surveillance cameras installed worldwide are still analog. As recently as the year 2014, two-thirds of the cameras sold were analog, and the share is even higher in the installed base of cameras.

Many analog cameras have been installed during recent years, and they will continue to be functional for the years to come. The average life expectancy of an analog surveillance camera is 5–7 years, but some last even longer. In many installations, the coaxial cable is the most expensive part of the installation. Once installed, there is therefore often limited incentive to recable with category 5 (CAT5) or category 6 (CAT6) cables to enable installation of a network camera. However, the recording device in an analog system—most often a digital video recorder (DVR) but sometimes even an older videocassette recorder—usually fails long before the camera component. This is where the video encoder comes into play.

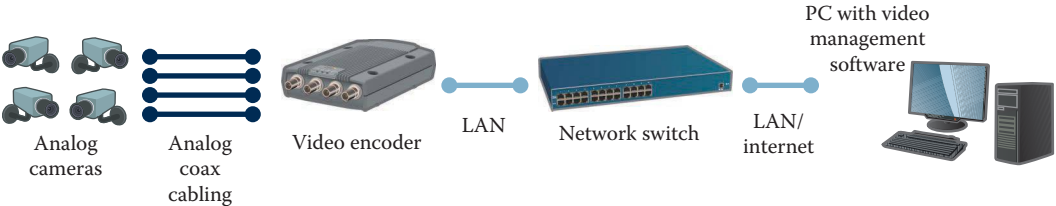
Video encoders allow security managers to keep their analog CCTV cameras while building a video surveillance system that provides the benefits of network video. If a video encoder is included in the system, analog cameras can be controlled and accessed over an IP network, such as a local area network or the internet, and old video recording equipment such as DVRs and monitors can be replaced with standard computer monitors and servers (see Figure 8.1).

Analog cameras of all types, such as fixed, indoor, outdoor, fixed dome, and pan, tilt, and zoom (PTZ), as well as specialty cameras such as covert, miniature, and microscope cameras, can be integrated and controlled in a network video system using video encoders.

The following sections give an overview of the components of a video encoder and the different types of video encoders available. The type of video encoder used depends on system configuration, camera count, camera types, and whether or not coax cabling is installed. This chapter concludes with an outline of a few best practices for encoders.

#### 8.1 THE COMPONENTS OF A VIDEO ENCODER

A video encoder connects to an analog camera through a coax cable and converts analog video signals into a compressed digital video stream that is transmitted over an IP-based network. The device is called a video encoder because it encodes video using a compression standard such as H.264 or Motion JPEG. Once the video is on a network, it is identical to a video stream coming from a network camera and is ready to be integrated into a network video system. A video encoder also often includes a serial port, which is commonly used for controlling the PTZ functionality of the connected analog camera.



**Figure 8.1** Analog cameras connected to a video encoder. This makes it possible to include the camera in a network video system.

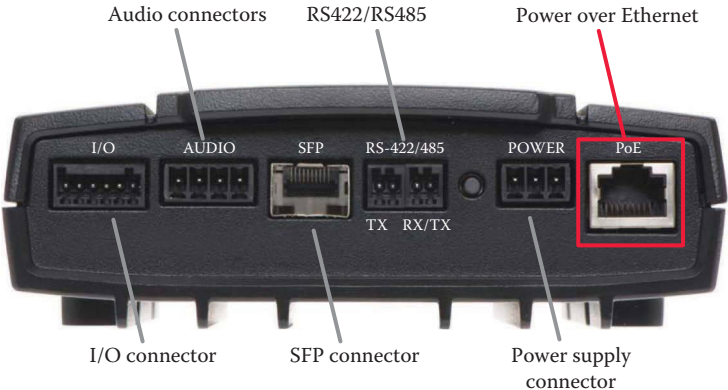
A video encoder also can offer many advanced functionalities, such as deinterlacing, video motion detection and other video analytics, alarm handling, one- or two-way audio support, and audio alarms. Third-party applications can also be uploaded to some video encoders to further enhance the system. Encoder models that have input/output ports can connect to external devices. This allows for control of other devices such as doors and lighting, and external sensors can be used to trigger an alarm event in the encoder.

Many video encoders also include Power over Ethernet (PoE) functionality (Figure 8.2). This enables the video encoder to receive power through the same cable used for data transmission and sometimes also to power the connected analog camera.

PoE can give substantial savings for the entire system because power cables can be excluded from the installation. In addition, if the server room is connected to an uninterruptible power supply, PoE enables the encoders to receive centralized backup power so they can continue to operate even in the event of a power failure.

The networking functionality is also very important, and it should include all the latest security and IP protocols. The video encoder's processor, which could be a general-purpose processor, a digital signal processor, or a purpose-built application-specific integrated circuit, determines the performance that is normally measured in frames per second per channel in the highest resolution, D1. The performance also depends on whether one chip is used per channel or if multiple channels share one chip. Today, the highest performing encoders can provide multiple individually configured video streams in full frame rate and still have headroom for running video analytics.

Some of today's video encoders can deliver up to 60 fps when connected to normal analog cameras. This is achieved by using powerful processors and new methods of deinterlacing. The benefit of this higher frame rate is a smoother video in high-motion scenes.



**Figure 8.2** An example of a *Power over Ethernet (PoE)*-enabled video encoder for connection to an analog camera. This image also shows connectors for I/O, pan, tilt, and zoom (PTZ) control (RS422/RS485), and audio, as well as connectors for a fiber-optic cable (small form-factor pluggable) and an external power supply that can be used if the network switch does not support PoE.

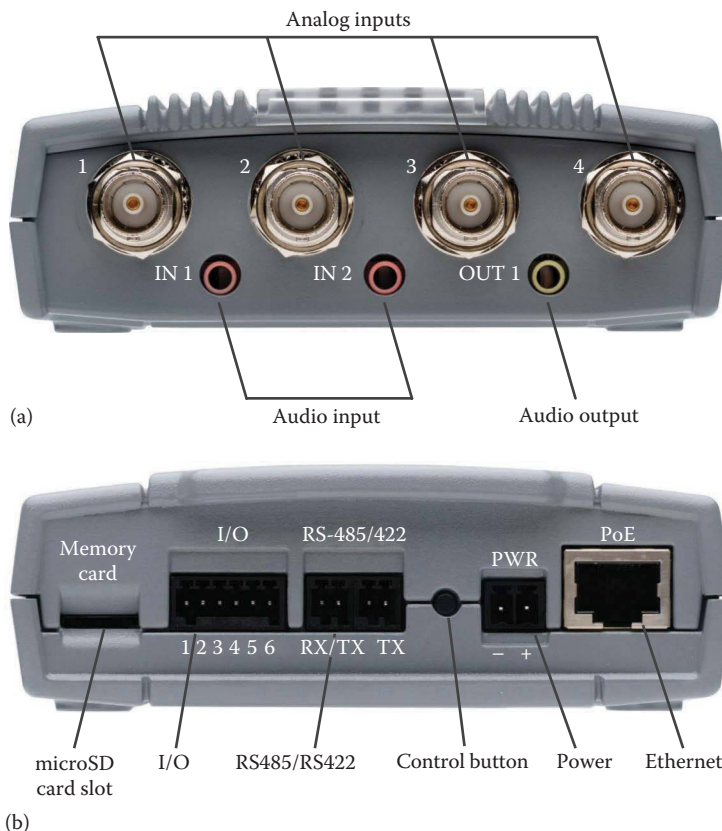
Many encoders have a memory card slot, which enables edge storage. This means that video can be recorded and stored locally. Local recording can be used in several ways. It could be the primary recording location to avoid the need for a server. It could also be used for redundancy. Even if the network fails, the video is still available because it is stored in the encoder. Finally, an operator can view live video in lower resolution or frame rate, while the encoder locally records full-quality video. This is a huge benefit in low-bandwidth environments because less data have to be sent over the network.

Today's video encoders have autosensing. In other words, they recognize if the incoming video is a PAL signal or an NTSC signal. The 75  $\Omega$  video termination for the video input can be enabled and disabled. In most cases, the best thing is to only enable termination in the last device in the video signal chain. For more information about PAL and NTSC, see Chapter 4.

Some video encoders are specifically designed for tough conditions and can be used in harsh environments, for example, where temperatures and vibration could normally be a problem. For long-distance network connectivity, some encoders have small form-factor pluggable (SFP) slots. This makes it possible to plug in an SFP module and connect the encoder to the network using a fiber-optic cable.

## 8.2 STAND-ALONE VIDEO ENCODERS

The most common video encoders are stand-alone versions that offer one or multiport (often 4 or 16) connections to analog cameras (Figure 8.3). Multiport video encoders have better cost efficiency, but sometimes the performance and flexibility can be limited. A multiport encoder is ideal



**Figure 8.3** Images showing the front (a) and back (b) of a typical stand-alone video encoder with four video inputs, I/O ports, SD card slot for local storage, audio, and serial ports (RS485/R422).



**Figure 8.4** A small, single-port video encoder positioned next to an analog camera in a camera housing.

in situations where, for example, there are a few analog cameras located in a remote facility or in a place far away from a central monitoring room. The video signals from the remote cameras can then share the same network connection on the video encoder, which dramatically reduces cabling costs.

In situations where investments have been made in analog cameras but coaxial cabling has not yet been installed, it is best to use and position stand-alone video encoders close to the analog cameras (Figure 8.4).

Placing the camera and video encoder together like this reduces installation costs. Because the video can travel over existing network cabling, there is no need to run new coaxial cables to a central location. It also eliminates the loss in image quality that would occur if video had to travel over long distances through coaxial cabling. A video encoder produces digital images, so the distance has no effect on the image quality. But with coaxial cables, the signal strength decreases the further the signals have to travel.

### 8.3 RACK-MOUNTED VIDEO ENCODERS

Most companies have a dedicated control room to gather equipment in one location so that operations can be efficiently monitored in a safe and secure environment. In a building containing a large number of analog cameras, this means that vast amounts of coax cabling run to the control room.

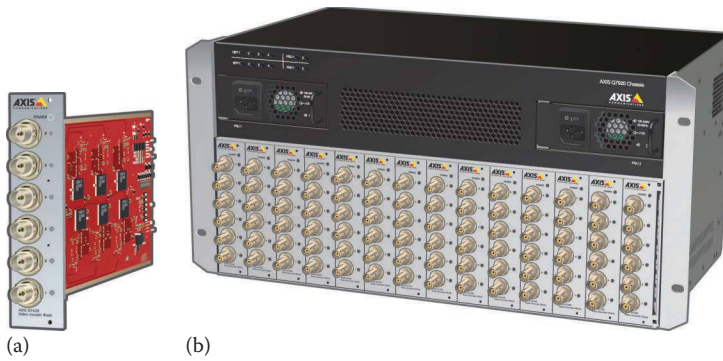
If all coax cabling has already been installed and is available from the central room, the installation would benefit from using a video encoder chassis with blade video encoders. A video encoder blade is basically a video encoder without a casing. A blade video encoder cannot function on its own. It must be mounted in a chassis.

A video encoder chassis allows a great number of video encoder blades to be mounted in a standardized rack, normally 19 in. wide, and managed centrally. Video encoder chassis offer functionalities such as an integrated network switch and hot swapping of blades, which means that blades can be removed or installed without having to turn off the power. Chassis have different heights depending on the number of slots for encoder blades they have (see Figure 8.5).

A chassis can provide network, serial communication, and I/O connectors, as well as a common power supply. A chassis is a high-density solution that can support up to 84 channels and save valuable rack space. Power and network redundancy can be used to make a high-density chassis very robust, secure, and reliable.

There is another popular video encoder version that usually has 16 channels and also fits into a 19 in. rack. But all its video channels are fixed mounted in a chassis instead of blades. While this solution lacks some of the redundancy and flexibility of a true rack-based system with blades, these encoders often have a lower cost per channel. Also, the DVRs they are replacing often have precisely 16 channels, and therefore the replacement process is very simple (Figure 8.6).





**Figure 8.5** An example of a six-channel encoder blade (a) and a video encoder rack that can support up to 84 channels (b).



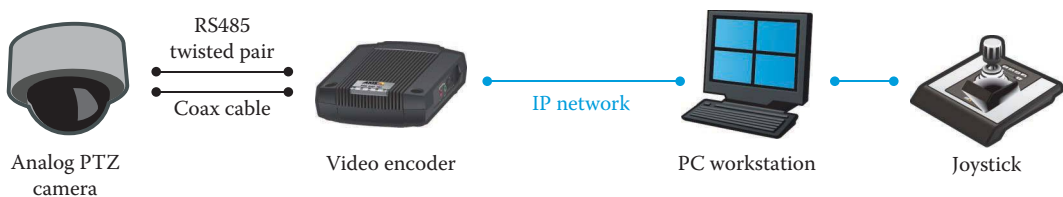
**Figure 8.6** Sixteen-channel video encoder in 19 in. rack-mountable chassis.

## 8.4 VIDEO ENCODERS WITH PTZ CAMERAS

The serial port (RS422/RS485) built into most video encoders is used to control the movement of analog PTZ cameras. In an analog CCTV system, separate serial wiring runs from the control board (with joystick and other control buttons) to the PTZ camera. But the separate wiring is not needed if a video encoder is placed close to the camera.

In a network video system, commands from the control board are carried over the same cable as the video and are forwarded by the video encoder through the serial port to the PTZ camera. Video encoders therefore enable control of PTZ functions over long distances using the internet. To control a specific PTZ camera, a driver must be uploaded to the video encoder. Many manufacturers of video encoders provide PTZ drivers for most PTZ cameras. A driver installed on a PC that runs video management software can also be used if the serial port is set up as a serial server that simply passes on the commands.

RS485 is most commonly used for controlling PTZ functions (Figure 8.7). One of the benefits of RS485 is the possibility to control multiple PTZ cameras using twisted-pair cables in a daisy-chain connection from one camera to the next. The maximum distance of an RS485 cable—without using a repeater—is 1220 m (4000 ft).



**Figure 8.7** Through the encoder's serial port (RS485), an analog pan, tilt, and zoom camera can be remotely controlled over an IP network.

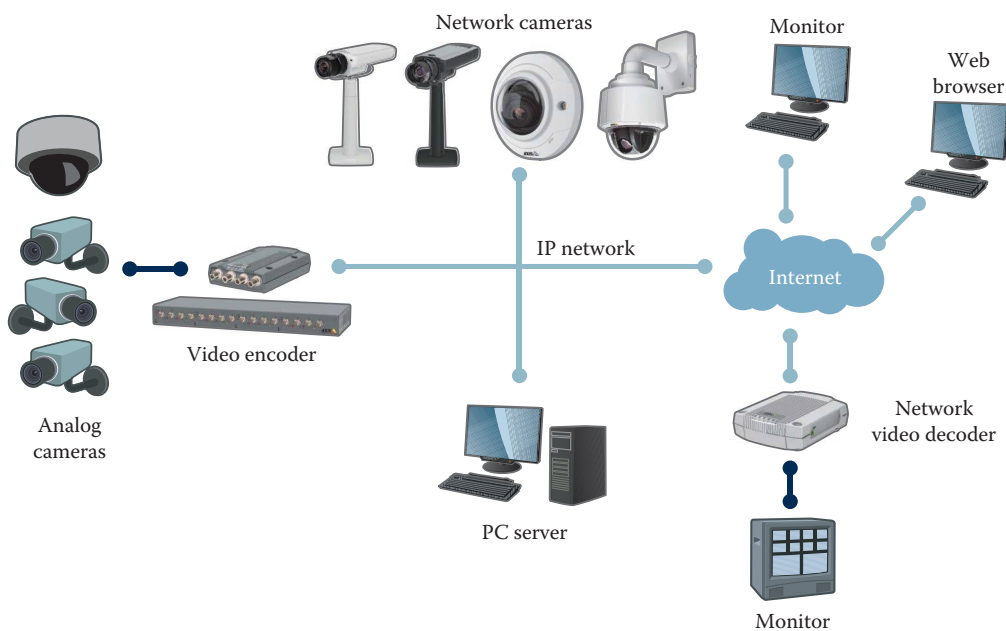
Some video encoders also allow for control of PTZ cameras through the coaxial cable connecting the video encoder and the analog camera. This is sometimes referred to as “up the coax.” If a compatible camera is used, no additional cabling is needed for PTZ control, which reduces installation costs.

## 8.5 VIDEO DECODER

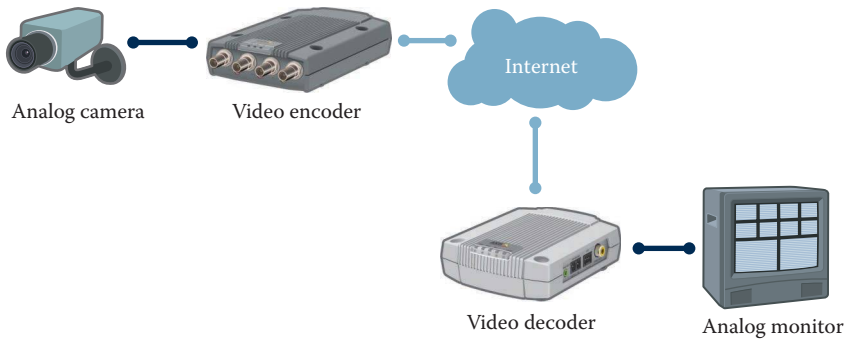
In some installations, there is a need to watch and listen to the network video and audio streams on existing monitoring equipment. When using a video decoder, the network video and audio streams are converted back to analog signals. These can then be connected to regular analog monitors and video switches. A typical case is a retail environment where the user might want to have traditional monitors in public spaces to demonstrate that the area is under video surveillance. A video decoder is used to connect such monitors to a network video stream, which comes from either a video encoder or a network camera.

Some video decoders have the ability to decode video from several cameras sequentially. This means that the decoder is decoding video from one camera for 5 seconds and then automatically changes to the second camera, then the third, and so on. This feature allows a guard to sit in front of a monitor and watch video, for example, from the five most important cameras (Figure 8.8). Video decoders can sometimes also show a split view from more than one camera on the monitor.

Another common application for video decoders is to use them in an analog-to-digital-to-analog setup for transporting video over long distances (Figure 8.9). Distance does not affect the video quality when images are sent in digital format. The downside is that there could be some level of latency, from 100 milliseconds to a few seconds, depending on the distance and the quality of the network between the endpoints.



**Figure 8.8** With a network video decoder, existing analog monitors can be used to show video and play audio from remotely located cameras—even those located in a different city.



**Figure 8.9** An encoder and decoder can be used to transport video over long distances from an analog camera to an analog monitor.

## 8.6 BEST PRACTICES

Video encoders offer a valuable solution to the challenge of migrating analog CCTV video to network video. Video encoders play a significant role, particularly in enterprise installations where large investments may have been made into analog cameras. Thanks to video encoders, the installation can be sustained longer and the investment protected.

It is easy to view video encoders as quite straightforward pieces of technology that are little more than analog-to-digital converters. In reality, however, the demands on video encoders are very high, and there are several considerations to make when selecting a video encoder.

Considerations to make include the following:

- *Image quality:* Can the video encoder deliver high-quality, deinterlaced digital video at a high frame rate? (For more information about deinterlacing, see Chapter 4.)
- *Resolution:* What resolutions can the video encoder offer?
- *Compression:* What compression standards does the video encoder support?
- *Performance:* How many channels that can deliver full frame rate at full resolution does the video encoder have? Can it send multiple simultaneous streams? Can the streams be configured individually?
- *Intelligent video support:* Does the encoder support a platform for adding intelligent video analytics from a third-party vendor?
- *Rack solution:* Are there rack-mounted versions of the video encoder available?
- *Density:* How many analog channels per chassis or per Ethernet channel can the system handle?
- *PTZ control:* Does the encoder support PTZ control over the coaxial cable?
- *Onboard storage:* Does the encoder support onboard storage using SD card memory for redundancy?
- *Audio:* Does the surveillance situation need support for audio?
- *External devices:* Is it possible to connect external devices to the encoder to create a more intelligent system?

Video encoders typically fall into the category of products that no one thinks about until something fails. Consequently, reliability and quality are key criteria for video encoders. Video encoders are advanced products that demand careful investigation when making a purchasing decision.



# Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

## CHAPTER 9

### Wired networks

Networks provide data exchange between servers and nodes in a computer system. In the early days of office and enterprise networking, many different technologies emerged, such as Token Ring, Banyan VINES, Ethernet, and Fiber Distributed Data Interface. Ethernet became the prevailing standardized technology for everything from home networking to large enterprise systems.

The following three chapters discuss different aspects of networking. This chapter provides an overview of wired networks, with a focus on Ethernet. Chapter 10 discusses wireless networks. Chapter 11 goes technical and focuses on networking technologies, including Open Systems Interconnection (OSI) layers, protocols, and network security. The purpose is to provide an overview in relation to network video, not to go deep into each aspect of networking technology. There are other books that aspire to do that.

#### 9.1 EVOLUTION OF ETHERNET

---

Bob Metcalfe at Xerox PARC documented his invention of Ethernet in 1973. During the remainder of the 1970s, Metcalfe and his colleagues continued to develop prototypes, published papers, and founded 3Com to commercialize Ethernet. In 1983, the Institute of Electrical and Electronics Engineers (IEEE) approved the IEEE-802.3 standard. This basic version of Ethernet enabled a data transfer rate of 10 Mbit/s (megabits per second), with individual nodes networked through a coaxial cable. Since then, the standard has improved continuously, and new transfer media achieves increasingly higher data transfer rates.

Today, Ethernet is based mainly on twisted-pair copper cables or fiber-optic cables (often simply called fiber). Coax cables can also supply Ethernet through Ethernet-over-coax converters. In smaller systems, Ethernet over power lines is a possible solution. Various manufacturers offer Ethernet components for building cost-effective networks. The number of networked nodes in a single network can range from two to several thousand. The data rates available depend on the transfer media and networking equipment used in each case. Some still install 10 Mbit/s Ethernet, but for network video this is not usually enough. The majority of networks get data rates that range from 100 to 10,000 Mbit/s, but enterprise networks may even get a data rate of up to 100 Gbit/s. Today's Ethernet networks easily provide the performance level required by the most demanding network video applications.

The Ethernet standard is available in many versions depending on the transfer medium used and the achievable data rate. The following subsections briefly describe the most important versions in terms of network video.

### 9.1.1 10 Mbit/s Ethernet

The 10BASE-T standard (802.3i) was released in 1990 and uses a twisted-pair cable and operates at 10 megabits per second (Mbit/s or Mbps). A twisted-pair cable is similar to an improved version of a telephone cable. It consists of four pairs of two twisted wires, which improves the electrical properties for data transfer (see Figure 9.1). A 10BASE-T system connects with RJ45 plugs and jacks and uses two of the four pairs of wires to transfer data. The maximum length of a cable segment is 100 m (328 ft).

Although 10 Mbit/s Ethernet still exists in legacy installations, it is not feasible for network video. Computers, switches, and other devices with a 10/100 interface support both 10 Mbit/s and Fast Ethernet. They automatically recognize and adjust to the current network speed.

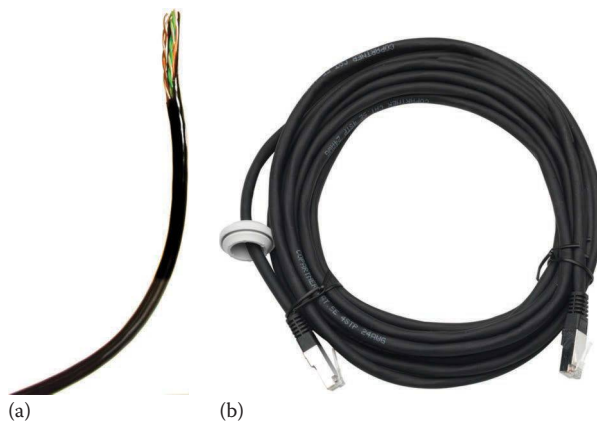
### 9.1.2 Fast Ethernet

The term *Fast Ethernet* refers to a 100 Mbit/s Ethernet network. Fast Ethernet was introduced with the 802.3u extension as 100BASE-T in 1995 and is described in the standard as a variation for twisted-pair cable (100BASE-TX) and glass fiber (100BASE-FX). 100BASE-TX provides backward compatibility with 10BASE-T and was the most popular Ethernet interface for a long time. While small businesses and residential markets may still use Fast Ethernet, most security networks have gigabit uplinks between their edge and core infrastructures.

### 9.1.3 Gigabit Ethernet

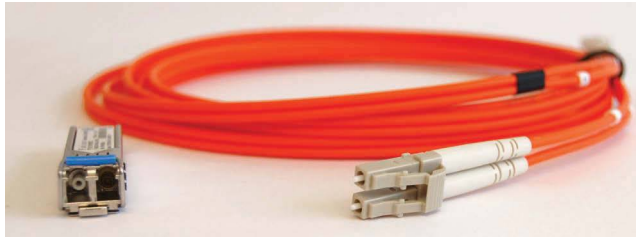
The third generation of Ethernet was specified in 1998 by the 802.3z extension. It includes specifications for optical fiber (1000BASE-SX, 1000BASE-LX) and short-distance copper cable (1000BASE-CX). In 1999, the IEEE updated the standard with the 802.3ab extension for twisted-pair cable (1000BASE-T). Gigabit Ethernet delivers a data rate of 1000 Mbit/s (1 Gbit/s, also sometimes written as 1 Gb/s or 1 Gbps). The main difference, when compared with 10BASE-T and 100BASE-TX, is that 1000BASE-T uses all four pairs of twisted wires in the cable to achieve the high data rates. Most gigabit interfaces are backward compatible with 10 Mbit/s and Fast Ethernet and are known as 10/100/1000 interfaces.

Various Gigabit Ethernet versions, such as 1000BASE-SX, 1000BASE-LX, 1000BASE-LX10, and 1000BASE-BX10, are available for use with fiber (see Figure 9.2) for transmission over longer distances and different wavelengths. For example, 1000BASE-SX works with multimode glass fibers (MMFs), has a wavelength of 770–830 nm, and permits cable lengths of up to 550 m (1804 ft). 1000BASE-LX has a wavelength of 1270 nm and permits cable lengths of up to 550 m (1804 ft) with multimode fibers (MMFs) or up to 5000 m (16,404 ft) with single-mode fibers (SMFs).



**Figure 9.1** An example of twisted-pair cable with four pairs of twisted wires (a) and RJ45 plug connectors and a protective gasket (b).





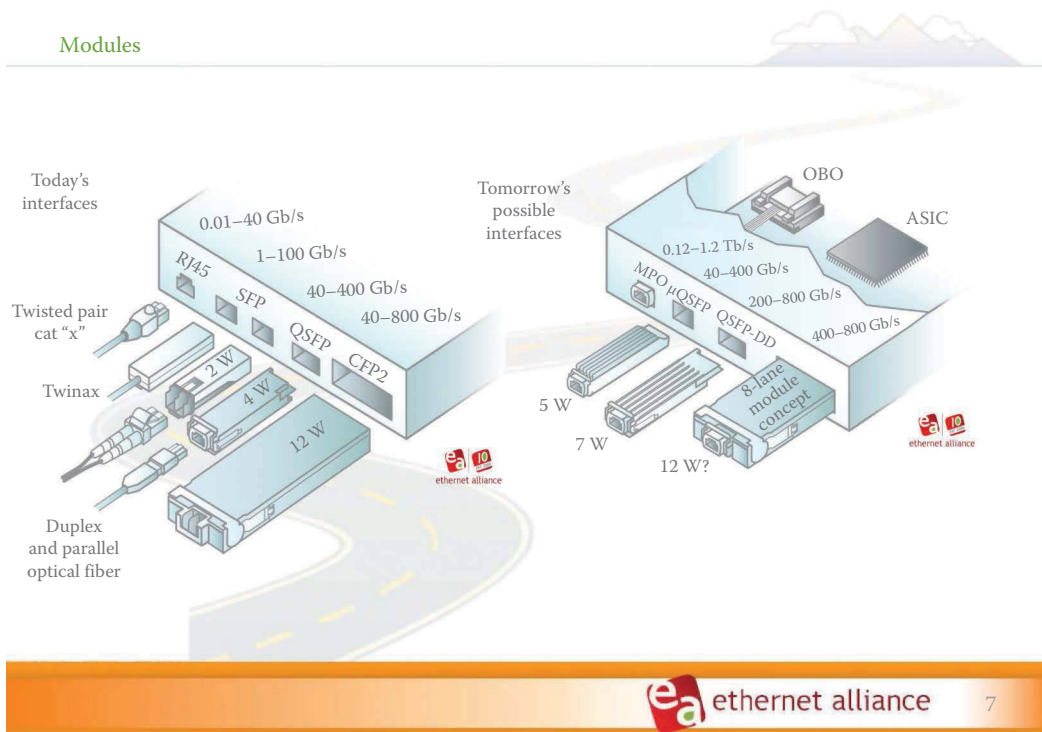
**Figure 9.2** An example of a small form-factor plug (SFP) module and a fiber-optic cable with SFP connectors. Fiber cables can bridge longer distances than twisted-pair copper cables. Backbone networks often use fiber cabling.

### 9.1.4 10 Gigabit Ethernet

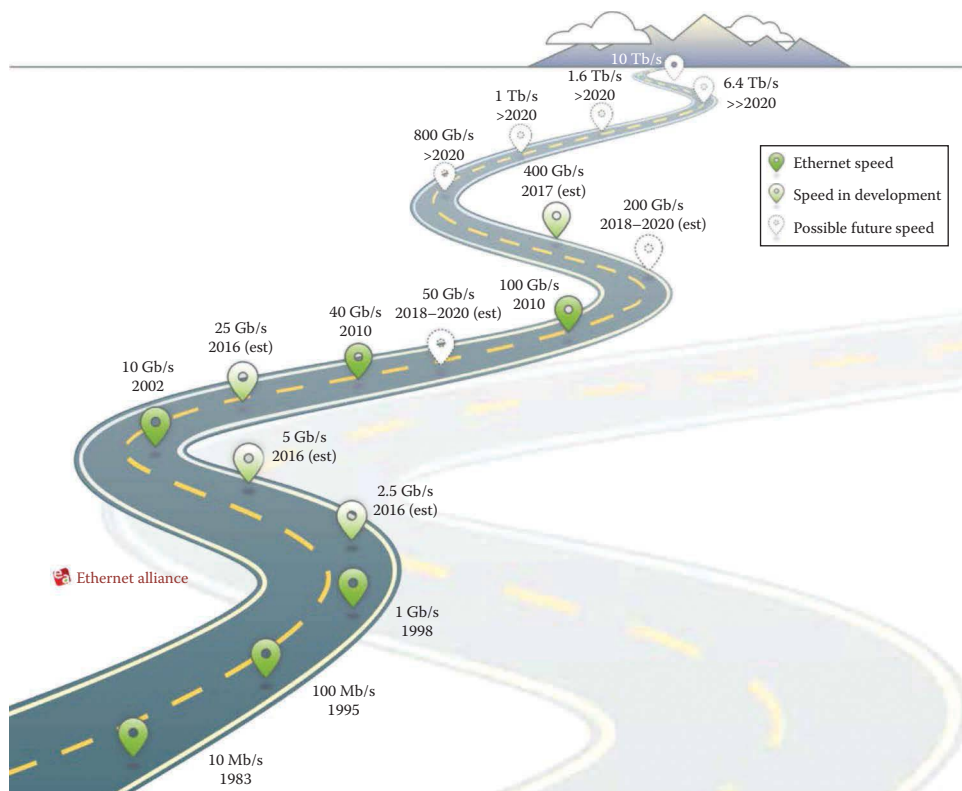
Gigabit Ethernet (also known as 10GE, 10GbE, or 10Gb Ethernet) delivers a data rate of 10 Gbit/s (10,000 Mbit/s). It was first defined in 2001 by the IEEE 802.3ae standard, which specified a number of Ethernet-over-fiber solutions. 10GBASE-LX4 and 10GBASE-SR can bridge distances up to 10,000 m (6.2 miles), and 10GBASE-ER can reach up to 40,000 m (24.9 miles). The 802.3an specification (10GBASE-T) was published in 2006 and permits data transfer of 10 Gbit/s through twisted-pair cable. It needs all four pairs of a high-quality cable (CAT6a or CAT7).

### 9.1.5 Future of Ethernet

The IEEE published 802.3ba in 2010, standardizing 40 and 100 Gbit/s Ethernet (also known as 40GbE and 100GbE). This standard specifies SMF (40GBASE-LR4, 100GBASE-LR4) or MMF (40GBASE-SR4, 100GBASE-SR10, -SR4) with quad small form-factor pluggable (QSFP) connectors (see Figure 9.3). As the term indicates, QSFP connectors support four channels



**Figure 9.3** An illustration of different Ethernet connectors. (Image courtesy of Ethernet Alliance, Beaverton, Oregon.)



**Figure 9.4** The 2016 Ethernet roadmap. (Image courtesy of Ethernet Alliance, Beaverton, Oregon.)

( $4 \times 10$  Gbit/s,  $4 \times 25$  Gbit/s). 40GbE is intended for local server connectivity and high-bandwidth applications such as video on demand, whereas 100GbE is intended for internet backbones and data centers. The video surveillance industry is not really using the 40 or 100 Gbit/s data rates yet.

In 2016, four new standards based on the 10GBASE-T technology for twisted-pair copper cables are becoming available. 25GBASE-T and 50GBASE-T are specified for 100 m of CAT5e cabling and 25GBASE-T and 40GBASE-T for 30 m of CAT8 cabling.

As we get closer to the year 2020 and beyond, we will see terabit links (see Figure 9.4). The idea is that 100 Gb/s lanes can be grouped into 10 and 16 lanes for 1 and 1.6 Tb/s links. However, investments have to be made and technology has to advance before anyone can implement such an infrastructure in a practical and economically sustainable way.

## 9.2 NETWORK TOPOLOGIES

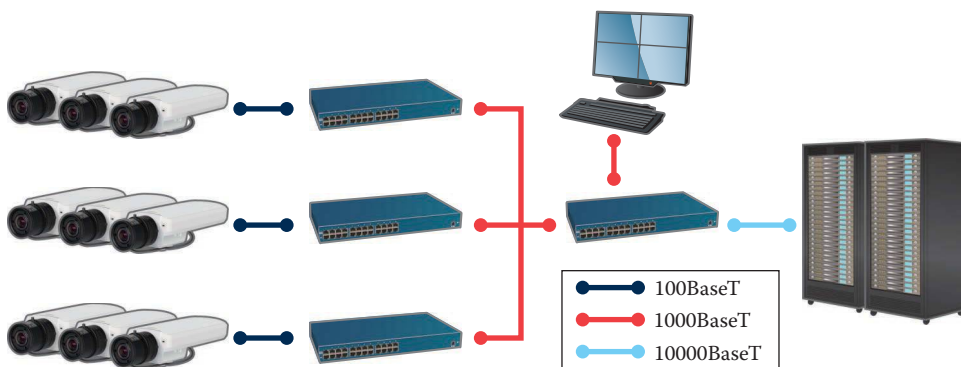
Networks can be built in different ways. The network configuration, also called the topology, describes how the individual nodes connect to the network. Most video surveillance networks today use a star topology.

In a star topology (see Figure 9.5), the individual nodes connect in a star formation through a central point, such as a switch. Twisted-pair or fiber cables form a point-to-point connection between the connected nodes and the central switch.

In larger network installations, multiple star topology networks connect in a hierarchy (see Figure 9.6). An uplink is used to make connections between the network switches. The central parts of the network that connect all the local star topologies are often referred to as the network backbone.



**Figure 9.5** In a star topology, all nodes connect to a central point.



**Figure 9.6** Larger networks are built in a hierarchy, with local star networks connected through a backbone.

## 9.3 NETWORK CABLING

In wired networks, all nodes must be connected through some type of network cable. Many different types are available, including twisted-pair or fiber cable. Each type of cable is available in many different versions. Twisted-pair cables are separated into different categories (CAT). This section outlines the most common types and versions.

### 9.3.1 Twisted-pair cables and RJ45

Twisted-pair cable is still the most common type of cable used in Ethernet networks. As mentioned earlier in this chapter, this cable has four pairs of twisted wires and uses RJ45 connectors. Wire pairs, which keep electromagnetic interference (EMI) low, transfer complementary signals. The maximum length of a twisted-pair cable for Ethernet is normally 100 m (328 ft).

Depending on the electrical properties of the twisted-pair cable, data transfer of up to 40 Gbit/s is achievable. In 10 and 100 Mbit/s Ethernet, only two of the four wire pairs are used for data

transfer—one pair of wires for sending data and another pair for receiving data. For 1,000 Mbit/s (Gbit) and 10,000 Mbit/s (10 Gbit) Ethernet, you need all four wire pairs. In some cases, the signals move simultaneously in both directions to reduce the frequencies.

### 9.3.2 Cable categories

Different data rates (measured in Mbit/s) are supported through different transfer frequencies, which are measured in megahertz (MHz). Different types of cables, referred to as categories or CAT, support the various transfer frequencies. The ISO/IEC-11801 standard defines the individual categories and specifies certain transfer properties for the twisted-pair cable, such as impedance, bandwidth, damping, and near-end cross talk. The categories relevant for Ethernet are from CAT3 to CAT8:

- *CAT3* is a twisted-pair cable for transfer frequencies up to 16 MHz, which corresponds to the requirement of 10 Mbit/s Ethernet (10BASE-T). Due to the limitation of 10 MHz or 10 Mbit/s, CAT3 should no longer be installed.
- *CAT5* is a twisted-pair cable for transfer frequencies up to 100 MHz, which corresponds to the requirement of Fast Ethernet (100BASE-TX). CAT5 exists in many installations today and is still used in some new installations.
- *CAT5e* is a twisted-pair cable for transfer frequencies up to 100 MHz and is built to the same specifications as a standard CAT5 cable. The *e* indicates extended inspection measurements and that the cable complies with the requirements for operating Gigabit Ethernet (1000BASE-T).
- *CAT6* was originally specified for a twisted-pair cable with transfer frequencies up to 250 MHz. To reduce signal noise, cross talk, and interference, it has thicker gauge wire and more pair twists per inch.
- *CAT6a* supports transfer frequencies of up to 500 or 625 MHz in special versions. It meets the requirements of 10GBASE-T and permits distances of 100 m. Unlike the majority of CAT6 cables, CAT6a cables are usually shielded. The shield reduces the risk of interference and therefore these cables are a better choice in industrial environments. CAT6e was introduced by manufacturers, but has never been a ratified standard.
- *CAT7*, also known as class F cabling, is a twisted-pair cable with transfer frequencies of up to 600 MHz. It is used in 10GBASE-T and 1000BASE-T networks. Its twisted-wire pairs are fully shielded. This is known as screened-foiled twisted-pair (S/FTP) wiring or, sometimes, as screen-shielded twisted-pair wiring. These cables successfully prevent cross talk and provide great noise resistance.
- *CAT7a* is the augmented version of CAT7 that has a frequency of 1000 MHz and can be used up to a maximum length of 100 m.
- *CAT8* meets the requirements for 25GBASE-T and 40GBASE-T. It is backward compatible with CAT6, uses RJ45 connectors, and has more or less the same diameter as CAT6a and CAT 7. With a frequency of 2 GHz it is four times faster than CAT6a but has a maximum length of 30 m. Therefore, it is more suitable for data centers than network video systems.

### 9.3.3 Twisted-pair cable types

The primary difference among the different categories explained earlier lies in whether or not the cable is shielded and if so, which type of shielding it has. The shield provides electromagnetic protection for the twisted pairs within the cable. Good shielding improves the performance of the cable, but it also increases the manufacturing cost.

To differentiate between the levels of shielding, the cables have different codes that follow a specific structure, XX/XXX. The code before the slash describes the cable screen, that is, the shielding that wraps around a group of pairs. The code after the slash describes the shielding for the individual pairs.

**Table 9.1** Twisted-pair cable types

Abbreviation	Description	Cable screen	Pair shield
U/UTP or UTP	No overall screen (U) and unshielded twisted pairs (UTPs). The wires are only covered with the standard plastic cover that protects them against physical damage.	None	None
F/UTP	An overall foil (F) screen and UTPs.	Foil	None
SF/UTP	An overall braid screen (S), an overall foil (F) screen, and UTPs. The overall braid screen helps prevent EMI.	Braid, foil	None
S/UTP	An overall braid screen (S) and UTPs. It is similar to the U/UTP and F/UTP cables and has an integrated wire that helps shield the cable.	Braid	None
F/FTP	An overall foil screen (F) and foil-screened twisted pairs (FTPs).	Foil	Foil
U/FTP	No overall screen (U) but foil-screened twisted pairs (FTPs).	None	Foil
S/FTP	An overall braid screen (S) and foil-shielded twisted pairs (FTPs). The foil shield around each individual pair helps prevent cross talk.	Braid	Foil
SF/FTP	An overall braid screen (S), an overall foil screen (F), and foil-shielded twisted pairs (FTPs)	Braid, foil	Foil

*Abbreviations:* TP, twisted pair; U, unshielded; F, foil shield; S, braid shield.

Table 9.1 describes the main types of twisted-pair cables.

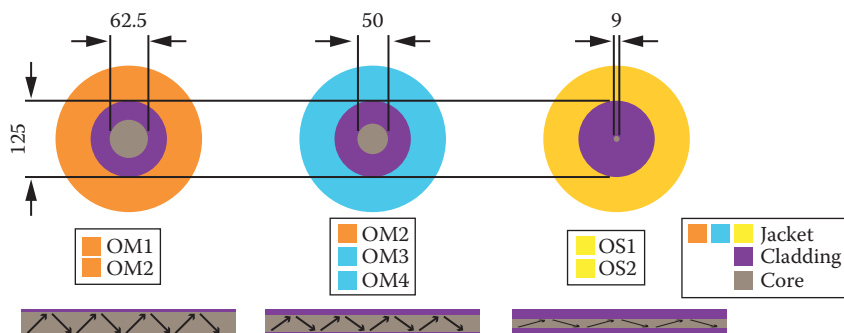
The term *shielded twisted-pair* (STP) often refers to cables with an overall screen and foil shields around each individual pair of copper wires. However, most twisted-pair cables with an overall screen or individual shields have at some point claimed the STP designation. Always make sure that the cables you intend to use have the right type of shielding for the installation. The higher the transfer frequency, the better the shielding needs to be. As mentioned in the beginning of the section, shields prevent EMI and help ensure that data transfer is error-free.

### 9.3.4 Fiber cable types

Fiber-optic cables are common in high-performance and backbone networks. Although more expensive, fiber offers several advantages. An obvious advantage is that the length of each cable segment can be much longer than a copper cable, which has a limit of 100 m (328 ft). Another advantage, especially in industrial environments, is that fiber is immune to EMI. In a network video application, fiber makes it possible to place a network camera further away from a building, for example, in a nonadjacent parking.

The core of fiber-optic cables is made of glass and light is the information carrier. Standard light-emitting diodes (LEDs) or special laser LEDs send and receive light at the two ends of the fiber. Depending on the glass fiber type, the LEDs use wavelengths of 850, 1300, or 1550 nm, and the light spreads through the glass fibers in different modes.

The two modes are MMF and SMF. The diameter of the cable is expressed as a ratio. MMF is typically 50/125 or 62.5/125  $\mu\text{m}$ , meaning that the core diameter is 50 or 62.5  $\mu\text{m}$  and the outer diameter of the cladding is 125  $\mu\text{m}$ . SMF is typically 9/125  $\mu\text{m}$ . As the light travels through the core, it bounces off the cladding. The bigger the diameter of the core, the more reflections and the longer time it takes for the light to reach the end of the cable. SMF, with its smaller diameter, is better for long distances and high bandwidth. MMF, with its larger diameter, is good for shorter distances. As with twisted-pair cables, fiber-optic cables are divided into different categories. They also have



**Figure 9.7** Typical dimensions of fiber-optic cables. OM1 and OM2 cables usually have orange jackets, OM3 and OM4 cables have aqua jackets, and OS1 and OS2 cables have yellow jackets. In single-mode cables, the light bounces fewer times than in multimode cables, which have larger cores.

**Table 9.2** Fiber-optic cables: categories and performance characteristics

				Minimum modal bandwidth (MHz × km)	
Fiber type	Core diameter (μm)	Wavelength (nm)	Attenuation (dB/km)	OFL <sup>a</sup>	EMB <sup>b</sup>
Multimode					
OM1	50/125 or 62.5/125	850	3.5	200	n/a
		1300	1.5	500	n/a
OM2	50/125 or 62.5/125	850	3.5	500	n/a
		1300	1.5	500	n/a
OM3	50/125	850	3.5	1500	2000
		1300	1.5	500	n/a
OM4	50/125	850	3.5	3500	4700
		1300	1.5	500	n/a
Single mode					
OS1	9/125 to 10/125	1310	1.0	n/a	n/a
		1550	1.0	n/a	n/a
OS2	9/125 to 10/125	1310	0.4	n/a	n/a
		1383	0.4	n/a	n/a
		1550	0.4	n/a	n/a

1 km is approximately 3281 ft or 1094 yd.

<sup>a</sup> Overfilled launch bandwidth.

<sup>b</sup> Effective modal bandwidth, also known as laser bandwidth.

different color jackets (see Figure 9.7). There are several standards that specify the various categories and performance categories of fiber, including EN 50173, ISO/IEC 11801, IEC 60793, and the TIA 492 series. Table 9.2 shows a summary.

### 9.3.4.1 Fiber connectors

When addressing network cabling issues such as distance or attenuation with optical fiber, it is important to match the fiber cable with the appropriate small form-factor pluggable (SFP) modules. Common standards include SX and LX. The SX standard is for shorter distances and uses 850 nm fiber cabling. The LX standard uses 1310 nm cabling to achieve longer distances. There are also data communication networks that use 1550 nm for extreme distances.

When connecting fiber cables, be sure to match the SFP or SFP+ modules with the connectors at the ends of the cable (see Figure 9.2). The electronics supplier should be able to recommend modules and cables that will fit the purpose and reach the target data rate.





**Figure 9.8** Using a media converter, you can connect any device that has an RJ45 jack (such as a network camera) to a fiber network. This particular example has two RJ45 jacks and two SFP jacks.



**Figure 9.9** A small form-factor pluggable module for a twisted-pair copper cable.

As the use of fiber optics in data networks has evolved, we have seen a variety of standards emerge. Early fiber-optic networks (with speeds of 10, 100, and 1000 Mbit/s) often used a larger form-factor connectors such as SC, ST, MTRJ, or SMA connectors where SC was the most common one in corporate networking environments. Because they are smaller and easier to use, LC connectors have replaced the SC connectors in popularity. LC connectors are often paired with SFP and SFP+ modules to connect the fiber-optic cables to switches, media converters, and other network devices.

Most network video products, such as network cameras, come with only a twisted-pair interface. Media converters, also known as fiber transceivers, make it possible to connect them with fiber cables anyway (see Figure 9.8).

There are also SFP and SFP+ modules for copper cables (see Figure 9.9). They provide a cost-effective way to connect devices over short distances, within racks, or across neighboring racks.

## 9.4 BASICS OF ETHERNET

The basic idea behind an Ethernet network is that the network provides a medium like the air through which all nodes can communicate. *Ether* is a word from the late fourteenth century with roots in old French *ether*, Latin *aether*, and Greek *ather*. It means “the upper regions of space,” therefore the name *Ether-net*. An early version of Ethernet was patented in a paper in 1977. In 1980, the IEEE formed the 802 committee to develop local area network (LAN) standards, and 3 years later, it formally approved the IEEE 802.3 standard. Though Ethernet has undergone substantial improvements, some of the basic elements still remain.

### 9.4.1 Media access control addresses

Media access control (MAC) addresses are used as source and destination addresses. A MAC address has a 48-bit address space, which allows for potentially  $2^{48}$  or 281,474,976,710,656 possible MAC addresses. These are unique addresses in hexadecimal format with a length of six bytes (e.g., 00-40-8C-18-32-78), where the first three bytes describe the manufacturer of the equipment



**Figure 9.10** Each networking device has a unique media access control address, which usually is printed on the product label so that you can identify the product even when it is offline. In this photo, it is shown as the serial number (S/N).

and the last three correspond to the serial number of the specific networking device (Figure 9.10). The manufacturer sets the MAC address and the user cannot change it. Each network device has a unique MAC address, which is often printed on the backside of the device. MAC addresses are always used when sending data from one device to another in a network, whether in a LAN or over the internet. For communication within a local network, MAC addressing may be all that is needed. In many cases, however, IP addresses are required in addition to MAC addresses. For more information about sending data packets, see Chapter 11.

### 9.4.2 Frames

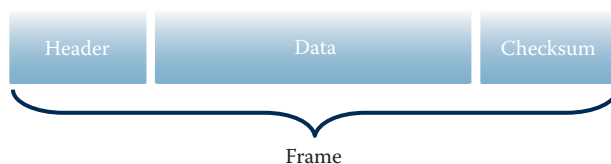
When data are transferred, they are packaged into frames. A frame consists of a header, the data, and a checksum through which the recipient can recognize transmission errors (Figure 9.11). The key information in the header is the destination and source addresses of the frame. The destination address specifies the node the frame should be sent to, and the source address specifies the node that sent the frame.

### 9.4.3 Half duplex and full duplex

Simplex means that data can be sent in only one direction. Half duplex means that data can be sent in two directions but only in one direction at a time. Full duplex means that data can be sent and received simultaneously, which in turn means better network performance.

Coaxial cables do not provide separate transmission and receiving channels and therefore only support half duplex. Twisted-pair cables and fiber-optic cables have separate transmission and receiving channels and therefore support full duplex.

Ethernet products with twisted-pair interfaces normally support several data rates in half-duplex or full-duplex mode. A twisted-pair interface can adjust itself automatically to the data rate and transfer mode through autonegotiation. Two nodes that are connected negotiate automatically to find the highest common data rate and transfer mode. If for some reason the negotiation fails, many devices allow users with administrator rights to set the transfer mode manually.



**Figure 9.11** A typical frame consists of the header, the data, and the checksum. This means that only part of the bits that are sent over a network consists of actual data. The remainder is called overhead.

## 9.5 NETWORKING EQUIPMENT

To network multiple nodes, you need equipment that can bridge between them. The most common network bridge is the network switch. Older networks may still use a hub instead. The network also uses other devices, such as a router, to connect to the internet.

### 9.5.1 Hubs

A hub, also called a repeater, is the simplest type of networking equipment (Figure 9.12). It works on the first layer of the OSI model (see Section 11.1.1). All nodes connect to the hub, forming a collision domain. Inside this collision domain, only one node can send data while all other nodes receive data at the same time. If another hub is connected, the collision domain is extended. If a hub receives data on one connection, it sends the data to all the other connections.

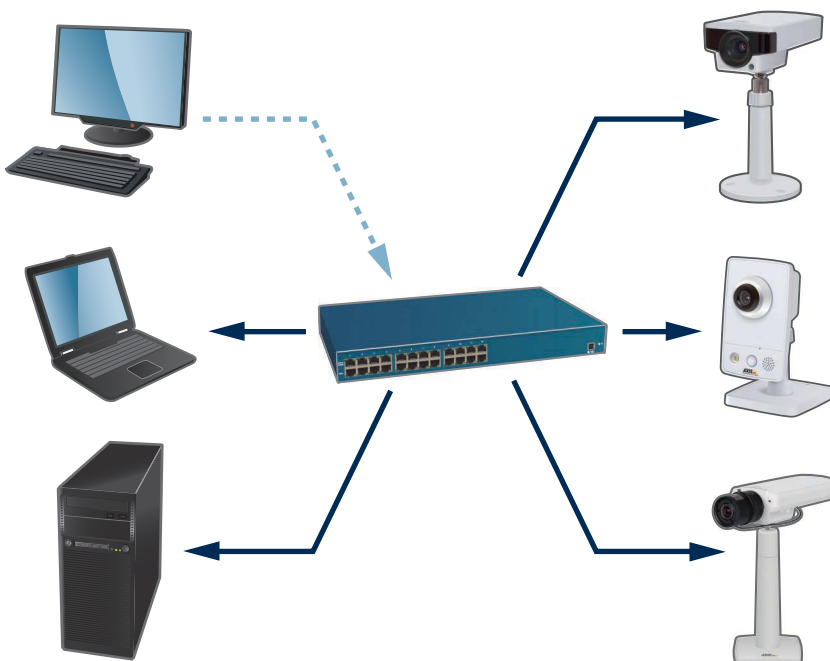
In a hub environment, all nodes operate in half-duplex mode. A classic hub can support only one data rate at a time, at 10 or 100 Mbit/s, meaning that all nodes in the network must support the same data rate. One exception here is the dual-speed hub, which supports two data rates. Hubs are rare in modern networks, which use switches instead.

### 9.5.2 Switches

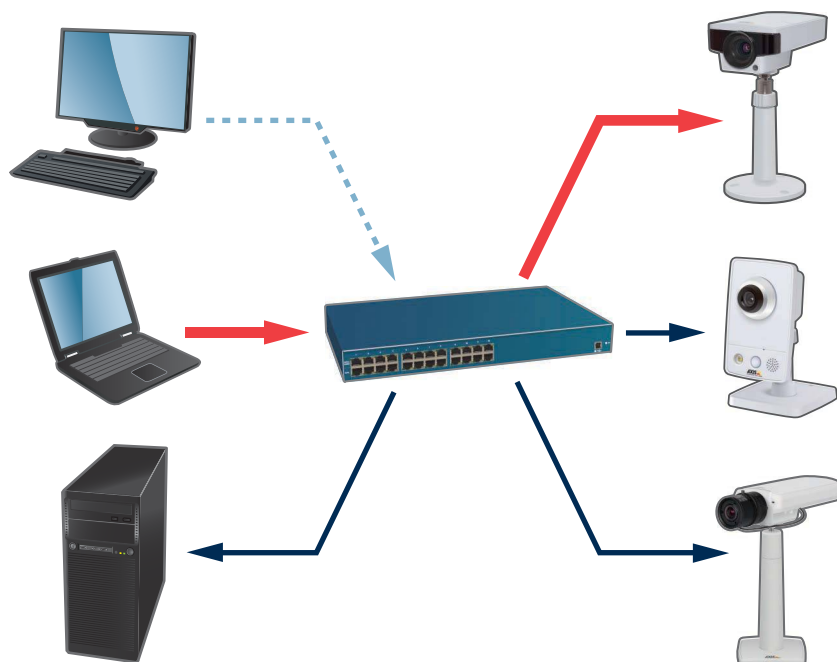
Network switches are more intelligent and can forward the network traffic in a much more efficient manner than hubs. Switches perform better and improve network security because they only send data to the devices that need to receive it, rather than broadcasting it as hubs do.

Whereas a hub operates on the first layer in the OSI model, a switch manages the data also on the second layer and, in some cases, the third layer in the OSI model. This is why some switches today are called Layer 3 switches. For more information about the OSI model, see Chapter 11.

The market has been confused about what the term *Layer 3 switch* means because different vendors use it to describe different functionality. A Layer 3 switch shares some of the functionality



**Figure 9.12** A hub is the simplest form of networking equipment. Several nodes can communicate with all other nodes connected to the hub.



**Figure 9.13** A network switch manages data transfer very efficiently because it directs the data traffic between the devices without affecting any other ports on the switch.

of a router (see Section 9.5.3). There are also other important Layer 3 features that are especially relevant to managing video. One is quality of service, which is required to manage the bandwidth. Another is Internet Group Management Protocol (IGMP) snooping, which is useful in multicasting networks. Some switches that have some Layer 3 functionality, such as IGMP snooping, are known as Layer 2+ switches. Chapter 11 provides more information about these technologies.

Data forwarding in a switch takes place through a learning process. The switch registers the address of the connected device. When the switch receives data, it forwards these data only to the port connected to the device with the right destination address (Figure 9.13). For more information about port forwarding, see Chapter 11.

Switches typically indicate their performance in per-port rates and in backplane or internal rates (both in bitrates and in packets per second). The port rates indicate the maximum rates on specific ports. This means that the speed of a switch (such as 1000 Mbit/s) is often the performance of each port.

A network switch normally supports different data rates simultaneously. The transfer rate and mode between a port on a switch and a connected device are normally determined through auto-negotiation. This means that data always travel at the highest common rate and through the best transfer mode. A switch also allows a connected device to function in full-duplex mode, that is, to send and receive data at the same time.

### 9.5.3 Routers

A network router is a device that routes information from one network to another. The most common job of a router is to connect a local network to the internet. Routers only forward the data packages that are supposed to be transmitted to another network. The main difference between switches and routers is that switches create networks, while routers link networks together.

A router can forward data between completely different network technologies, thus creating a larger interconnected network. Routers are sometimes referred to as gateways as, in reality, this is where

they are located that is, where two or more networks connect. Modern integrated routers typically include a multiport Ethernet switch, a Network Address Translation (NAT), and a Dynamic Host Configuration Protocol server.

### 9.5.4 Firewalls

Firewalls prevent unauthorized access to or from private networks. Firewalls can be implemented in both hardware and software or a combination of both. The most typical task for a firewall is to prevent unauthorized internet users from accessing private networks connected to the internet. Messages entering or leaving the internet pass through the firewall, which examines each message and blocks those that do not meet the specified security criteria.

### 9.5.5 Bridges

A network bridge is a device or software that connects different physical networks together while filtering the data traffic between them to provide a measure of network traffic segregation. A bridge makes packet-forwarding decisions based on the MAC address of the packets' destination or destinations. It forwards packets to the network port where the particular MAC address resides.

Layer 2 switches are actually multiport bridges that make network traffic forwarding decisions based on MAC addresses. A router, on the other hand, makes packet-forwarding decisions based on IP addresses. A Layer 3 switch is, in effect, a multiport router. A network gateway makes forwarding decisions based on either MAC or IP addressing, but adds network translation functions such as NAT, possibly provides firewall function, and acts as a conversion point between network connectivity types such as cellular and wired.

### 9.5.6 Internet connections

To connect to the internet, an internet service provider (ISP) must establish the network connection.

People use the terms *upstream* and *downstream* when they describe the transfer rate between their devices and the internet. Upstream describes the speed at which they can upload data from the device to the internet, such as when network camera sends video to storage device in the cloud. Downstream describes the speed at which they can download files or stream video to a viewing device.

Considering that most internet users consume data rather than produce it, the download speed is usually the most important. In a network video system, however, where network cameras are located at remote sites, the upstream speed is more relevant. This is because the network cameras need to upload data (video) to the internet.

Common internet connection technologies include the following:

- **Cable:** A cable connection uses an existing cable TV connection, overlaying the data communication on the cable used for TV.
- **T1:** A T1 connection provides a bandwidth of 1.544 Mbit/s for both upstream and downstream data transfer. Several T1 connections can be pooled by a device that provides 3, 6, and 9 Mbit/s connections.
- **Fiber:** A fiber connection means that a fiber-optic cable connects directly into an ISP's network, providing data transfer speeds of typically 100 Mbit/s or more, in some cases up to 100 Gbit/s, to the internet.

The DSL, ADSL, and SDSL technologies are irrelevant in today's video surveillance networks.

## 9.6 POWER OVER ETHERNET

Power over Ethernet (PoE) provides the option of supplying devices with power and Ethernet through their network cable. The whole idea stems from older telephone systems where the telephone line was both a means for communication and mains power. In the early days of IP telephony,

the same functionality was mimicked using Ethernet cabling. Today, PoE is widespread and is used to power IP phones, wireless devices, and network cameras in Ethernet networks.

The primary benefit of using PoE is the inherent cost reduction. Depending on the camera location, not having to install a designated power cable can save up to several hundred dollars. It also makes it easier to move a camera to a new location or add cameras to a video surveillance system.

In addition, PoE can make a video system more secure. A video surveillance system with PoE can be powered from a server room, which is often backed up with an uninterruptible power supply. This means that the video surveillance system can be operational even during a power outage.

9.6.1 802.3af and 802.3at standards

Most PoE devices today conform to the IEEE 802.3af standard or the updated version, the IEEE 802.3at standard, which were published in 2003 and 2009, respectively. The original 802.3af standard supports CAT3 cables for low-power devices but requires CAT5 cables for higher-power devices and to ensure that data transfer is stable. The 802.3 at standard requires CAT5 cables or higher.

In these standards, the device that supplies the power is referred to as the power sourcing equipment (PSE). The PSE functionality can be built into a network switch or provided by a midspan (see Section 9.6.2). The device that receives the power is referred to as a powered device (PD). The functionality is normally built into a network device such as a network camera or provided in a stand-alone splitter (see Section 9.6.2).

Backward compatibility to non-PoE-compatible network devices is guaranteed. The standard includes a method for automatically determining whether a device supports PoE, and only when that is confirmed will power be supplied to the device. This also means that if the Ethernet cable is not connected to a PoE device, it will not supply power, even if connected to a PoE switch. This eliminates the risk of electrical shocks when installing or rewiring a network.

In a twisted-pair cable, there are four pairs of twisted wires. PoE can use either the two unused wire pairs or overlay the current on the pairs used for data transmission. As a rule, switches with built-in PoE supply electricity through the two pairs of wire used for transferring data, whereas midspans normally use the two unused pairs. A PD supports both options. A Type 2 Class 4 PD supports data transmission over four pairs.

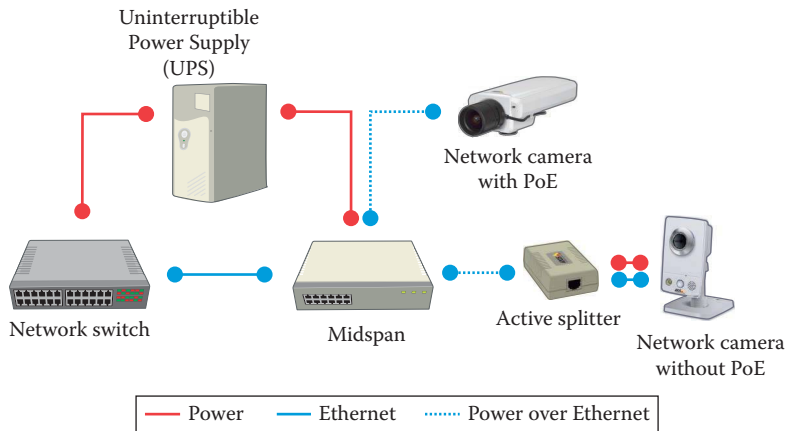
802.3af specified a PSE to deliver a voltage of 48 V DC at a maximum power of 15.4 W per port. Considering the power loss that takes place on a twisted-pair cable, only 12.95 W are guaranteed for a PD. For 802.3.at, the specification changed to 44–57 V DC over two pairs for Type 1 PSEs and 50–57 V DC over four pairs for Type 2 PSEs. Table 9.3 provides the various classes and power ranges according to the 802.3af and 802.3at standards. In the 802.3af standard, Class 4 was reserved for future use.

Some PSEs only deliver a certain amount of power. On a 48-port switch with 500 W, this would mean 10 W per port if all ports are connected to devices that use Type 1 PoE. Unless the PDs support power identification, the full 15.4 W must be reserved for each PoE port, which means a switch with 500 W can supply power on only 32 of the 48 ports. However, if all devices let the switch know that they are Class 2 devices, the 500 W will be enough to supply power to all 48 ports.

Table 9.3 Power over Ethernet classes and their power ranges

IEEE 802.3af PoE	IEEE 802.3at PoE+	Power at PSE (W)	PD power (W)
Class 0	Type 1 Class 0	15.4	12.95/13.0
Class 1	Type 1 Class 1	4.0	3.84
Class 2	Type 1 Class 2	7.0	6.49
Class 3	Type 1 Class 3	15.4	12.95/13.0
	Type 2 Class 4	30.0	25.5





**Figure 9.14** Midspans and splitters deliver Power over Ethernet functionality to existing systems.

Most fixed network cameras can receive power through PoE and are normally identified as Class 1, Class 2, or Class 3 devices. Pan, tilt, and zoom and other cameras with motor control are usually Type 2 Class 4 devices (PoE+). Cameras with heaters and fans usually require even more power, typically 60 W. Therefore, manufacturers have developed their own standards with names such as High PoE and UPoE.

### 9.6.2 Midspans and splitters

With PoE, two new network devices were introduced: the midspan and the splitter, also known as an active splitter (Figure 9.14). Both devices enable an existing network to support PoE.

The midspan, which adds power to an Ethernet cable, is placed between a network switch and PDs. Midspans with 1, 6, 12, 24, or 48 ports that support the IEEE 802.3af/802.3at Type 1 standard are readily available.

A splitter, or active splitter, splits the power and data in an Ethernet cable into two separate cables, which can then connect to a device that has no built-in support for PoE. Because the 802.3af standard supplies only 48 V DC, another function of the splitter is to step down the voltage to the appropriate level for the device, often by 12 or 5 V.

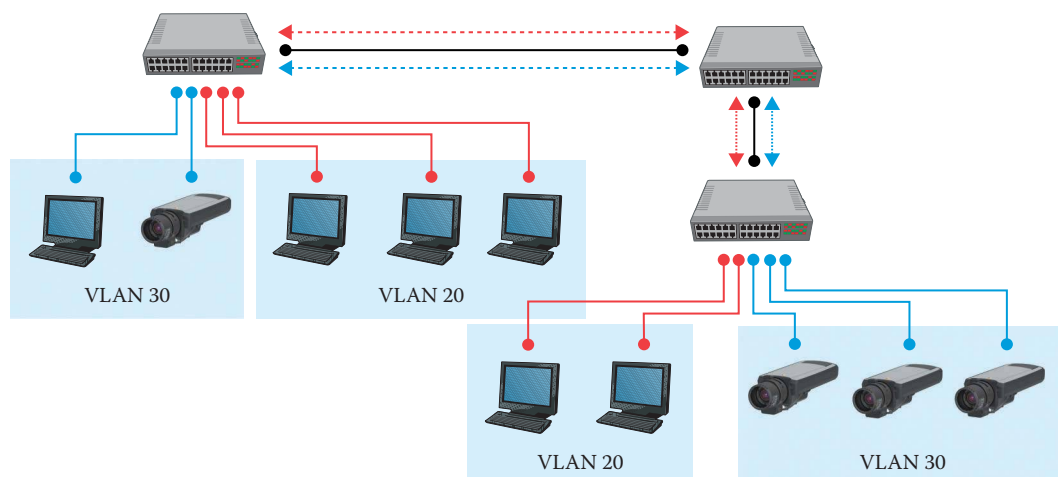
## 9.7 VIRTUAL LOCAL AREA NETWORKS

When designing a network video system, you often want to keep the network separate from other networks for both security and performance reasons. At first glance, the straightforward approach is to build a separate network. However, the cost of purchasing, installing, and maintaining the network would probably be higher than if you set up a virtual LAN (VLAN). This is a technology for virtually segmenting networks, a functionality that most network switches support. It can be achieved by dividing network users into logical groups. Only users in a specific group are capable of exchanging data or accessing certain resources on the network. If a network video system is segmented into a VLAN, only the servers located on that VLAN can access the network cameras (Figure 9.15).

The primary protocol for VLANs is IEEE 802.1Q, which tags each frame or packet with extra bytes to indicate which virtual network the packet belongs to. Before this standard became available, several manufacturers used proprietary VLAN protocols. VLANs operate on Layer 2 of the OSI model; see Chapter 10.

The four ways to assign VLAN memberships are as follows:

1. *Port based*: Each physical port on a switch is configured as part of either a VLAN or a LAN.
2. *MAC based*: Each MAC address that is part of a VLAN is listed in the switch.



**Figure 9.15** In this illustration, virtual local area networks (VLANs) are set up over several switches. First, each of the two different local area networks is segmented into VLAN 20 and VLAN 30. The links between the switches transport data from different VLANs. Only members of the same VLAN are able to exchange data, either within the same network or over different networks. VLANs can be used to separate a video network from an office network.

3. *Protocol based:* Layer 3 data are used to determine which VLAN a frame belongs to. For example, AppleTalk® is one VLAN and IP is another.
4. *Authentication based:* Devices are placed in a VLAN based on 802.1X authentication (see Chapter 11).

## 9.8 BEST PRACTICES

Network bandwidth was a scarce resource in the early days of network video. Although today's networks have less problems with bandwidth availability, there are several important aspects to consider when designing a surveillance video network:

- *Is there existing cabling?* If so, this can save a lot of money. Make sure that the cabling is of appropriate quality for the network speed and PoE class.
- *Does the system need new cabling?* If so, what category cable is appropriate? Is fiber an option? When looking at the cost of installing a cable, look at the cost of labor versus the cost of the cable itself. The cable is normally a smaller portion of the total cost. Choose as good a cable as possible to future-proof the installation. To ensure performance and PoE compatibility, use CAT5e or CAT6. Use CAT6 or CA7 cables for backbone networks.
- *What distance does the cables need to cover?* If the distance is greater than 100 m (328 ft), a fiber cable may be the best solution.
- *Is it possible to use PoE?* PoE provides huge savings, so make sure to use it for as many devices as possible. If using PoE, make sure that the power available in the switch or midspan is sufficient for the nodes and that the nodes support power classification.
- *What are the bandwidth requirements?* Make sure that the switches provide more than the minimum required bandwidth. To future-proof a network, design it so that the original deployment only needs 50%–70% of the capacity.

- *Is there a need for a wide area network (WAN) connection?* WAN bandwidth is normally limited. Design the system so that it does not overload the WAN bandwidth.
- *Is there a need for a VLAN or separate network?* VLANs normally provide better and more cost-effective solutions than separate networks.

A good rule of thumb is to build a network with greater capacity than is currently required. Because more and more applications run over networks today, the need for higher speeds and better network performance will continue to increase. Whereas it is easy to replace network switches after a few years, cabling is normally much more difficult to replace.



# Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

## CHAPTER 10

# Wireless networks

The use of wireless networks and devices has increased tremendously over the last few years. This process is likely to continue and accelerate even further, as billions of new devices are destined for connection to the “Internet of Things” in the near future. New products, increased data transfer speeds, better coverage, and new security technologies are giving us more opportunities to exchange information at lower cost and to greater advantage.

For video surveillance applications, wireless technology presents an interesting alternative to wired networks. In recent years, IP-based video applications have benefited from improved security and more plentiful bandwidth. Installing a wireless camera in a parking lot or a hard-to-reach location can often be cheaper and easier than installing cabling. Wireless also offers a way to deploy cameras over a large area quickly and efficiently, especially in urban surveillance applications.

Wireless networks also open up for mobility. A device such as a network camera can move around freely within the coverage range of a wireless network. A camera in a bus or train can also be accessed live from a remote location. It is relatively simple to extend the network by using various range extenders or repeater devices. In older buildings, which can be protected by heritage legislation, wireless networks may actually be the only alternative where the installation of standard network cables is prohibited.

A network camera with built-in wireless support can be integrated into an existing wireless network directly (Figure 10.1). If the camera lacks wireless support, a wireless bridge can be used to provide this functionality (Figure 10.2).

Wireless networks come in many different forms. A wireless local area network (WLAN) is predominantly based on the IEEE 802.11 standard. Other standards, as well as proprietary technologies, are also available, some of which are of interest for video surveillance applications. Wireless networks can be designed in different ways, as point-to-point, point-to-multipoint, and mesh networks. Different wireless networks operate in different frequency ranges, called spectra (or spectrums). One of the benefits of the 802.11 wireless standards is that they all operate in a license-free spectrum, meaning that there are no licensing fees associated with setting up and operating the network.

### 10.1 BASICS OF WIRELESS NETWORKS

---

Wireless networks involve the transmission of electromagnetic waves through the air. Ever since the Italian inventor Marconi conducted the first wireless transmission of Morse code in 1895, wireless technology has played an increasingly important role in people’s lives. The basic technology



**Figure 10.1** A network camera with built-in wireless networking.



**Figure 10.2** Using a wireless bridge, any network camera can connect to a wireless network.

remains the same, but the many developments over the past hundred years have brought about AM/FM radio, broadcast TV, mobile telephony, Bluetooth® wireless technology, and many other technologies.

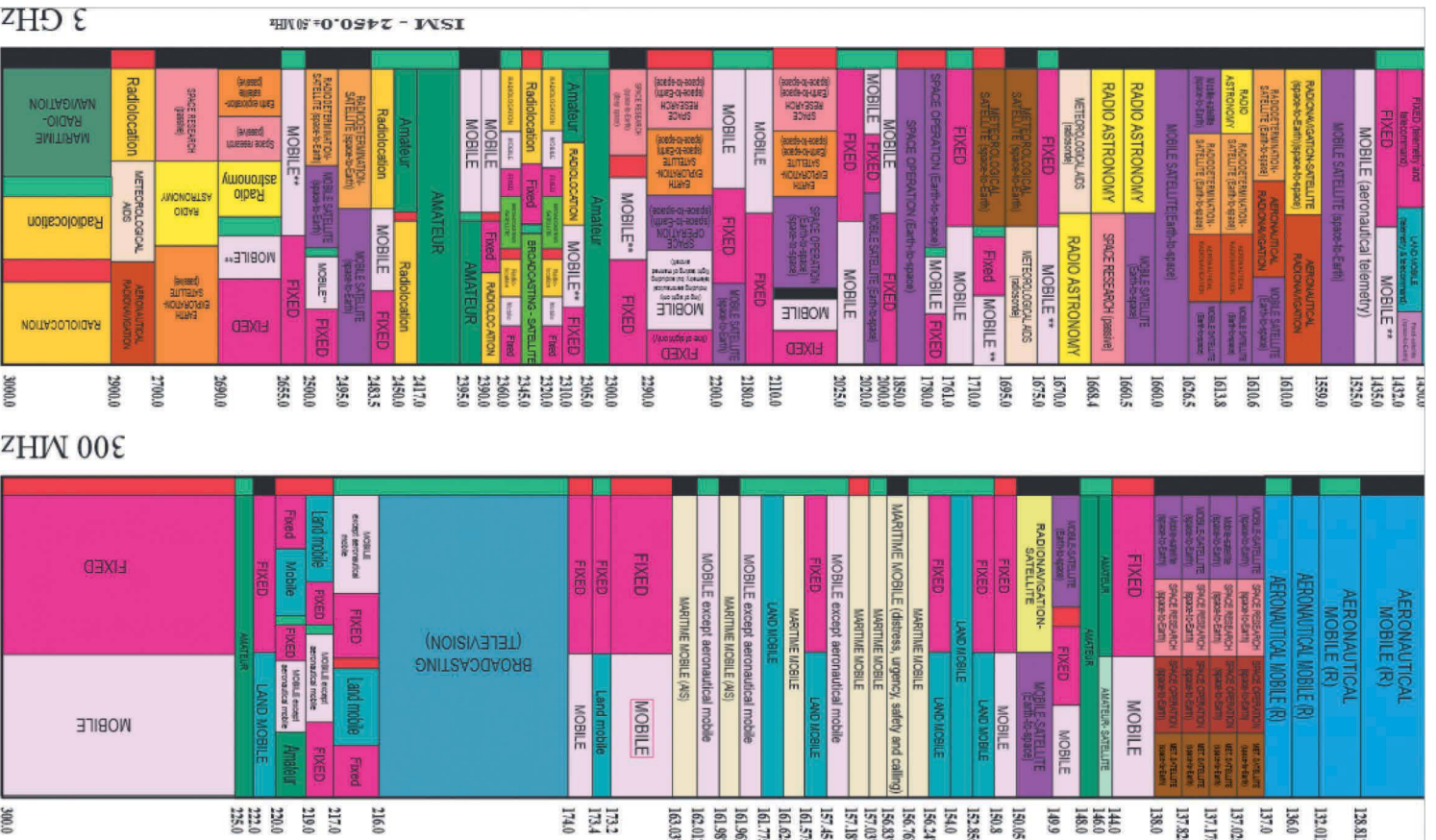
### 10.1.1 Wireless spectrum

Wireless communication is transmitted at certain frequencies measured in hertz (Hz), which is the unit for electromagnetic cycles per second. By using different frequencies, two or more different wireless systems can communicate at the same time, even within the same area. This is why, for example, a person can make a call on a cell phone while at the same time listening to an FM radio—these two systems are using different frequencies. If they were using the same frequency, there would be interference between them.

The risk of interference is one of the reasons why we have regulatory bodies to manage radio frequencies. Some frequencies are in the licensed spectrum, which means that governmental bodies such as the Federal Communications Commission must approve the use of certain frequency bands. Other parts of the spectrum are license-free, meaning that anyone can use them without prior approval and at no cost. See Chapter 17 for more on electromagnetic compatibility.

The lower the frequency, the greater the range of the radio signal. On the other hand, higher frequencies enable greater amounts of data throughput (bandwidth). The division and allocation of frequency ranges for different uses varies around the world. Figure 10.3 shows part of the spectra allocation in the United States. Most of the wireless spectrum from 300 MHz to 3 GHz is today already allocated. Applications such as FM radio (88–110 MHz) occupy the spectrum below 300 MHz.





**Figure 10.3** Part of the table of frequency allocations for the United States. (Image courtesy of U.S. Department of Commerce, National Telecommunications and Information Administration, Office of Spectrum Management, Washington, DC.)

## 10.1.2 Signal strength

There are different ways to measure signal strength in a wireless system. One way is to examine the amount of radio-frequency (RF) energy transmitted. The more energy transmitted, the greater the range of the signal. The unit for energy is watts (W), and in most wireless systems, the energy is in the milliwatt (mW) range. For example, in a typical 802.11 system, the maximum energy is 100 mW at the wireless access point. At higher energy levels, the range of the radio signal will be greater, but so will the risk of interference.

Although it provides an absolute value, measuring RF energy in mW is not always convenient. This is because a signal's strength does not fade in a linear fashion, but instead inversely, as the square of the distance. So if the signal level at a certain distance from a wireless access point is 100, the level at twice that distance will be 25; that is, it will have been reduced by a factor of 4. The fact that exponential measurements are involved in measuring signal strengths is one reason why the use of a logarithmic scale was developed as an alternative way of representing RF power. The dBm (dB milliwatt) is a logarithmic measurement of signal strength, and dBm values convert exactly and directly to and from mW, according to the following formula:

$$P_{dBm} = 10_{\log_{10}} (P_{mW})$$

Where  $P$  is the signal strength. Therefore, a signal strength of 100 mW = 20 dBm. Halving the signal strength in mW, reduces the signal strength by 3 dBm.

Other ways of measuring the quality of a signal are relative (i.e., not absolute). One such method is known as the receive signal strength indicator (RSSI) value. It is measured through a mechanism defined by the IEEE 802.11 standard. Here, RF energy is measured using the circuitry on a wireless network interface card and is then represented by a value in the numerical range 0–255. Different vendors interpret the RSSI value in different ways, and the *RSSI\_Max* value may also be selected arbitrarily within this range.

## 10.1.3 Antennas

The antenna is an important part of a wireless system, and a better antenna will increase the range of the wireless connection. The performance, or gain, of an antenna is measured in decibel isotropic (dBi), which describes the gain a given antenna has over a theoretical isotropic (point source) antenna, normally in the range 1–20 dBi. The effective radiated power (ERP) is defined as the effective power of the transmitter antenna. It is equal to the sum of the antenna gain (in dBi) plus the power (in dBm) into the antenna. For example, a 12 dBi gain antenna fed directly with 15 dBm of power has an ERP of

$$15 \text{ dBm} + 12 \text{ dBi} = 27 \text{ dBm (500 mW)}$$

A longer antenna normally has a higher gain. The appropriate length of an antenna depends on the frequency; to enable the highest efficiency, the antenna length should be a multiple of the wavelength.

The two main types of antenna are omnidirectional and directional (Figure 10.4). An omnidirectional antenna spreads the electromagnetic waves in all directions, making it appropriate for multi-point networks where wide coverage is required. A directional antenna focuses the electromagnetic waves in a single direction, enabling the radio signal to reach further, but only in that direction. Directional antennas are appropriate for point-to-point applications and come in different types, such as Yagi antennas, patch antennas, and parabolic grid antennas. Many have a gain between 10 and 30 dBi.



**Figure 10.4** Communication towers equipped with omnidirectional antennas (stick shaped) and directional antennas (rounded).

### 10.1.4 Radio wave propagation

Although open air is best for transmitting radio signals, radio signals do not require free line of sight for transmission. This means that radio systems are so-called non-line-of-sight systems, as opposed to optical transmission systems, where the receiving and the transmitting ends must be able to “see” each other. Some systems have better ranges, some have the ability to pass through certain materials, and some have both, as explained in the following.

A wireless signal propagates best through open air, but technically, it can pass through most other media, such as the wood and concrete present in buildings. Structures containing large amounts of metal present difficulties for radio signals, as signals reflect off metals instead of penetrating them. A building containing or covered by a metal with high conductivity such as copper will be virtually impenetrable for a wireless signal. Such a building or room is called a Faraday cage, after the scientist Michael Faraday.

Radio signals at lower frequencies and longer wavelengths propagate further than signals at higher wavelengths. Lower radio frequencies also have a better ability to penetrate materials such as wood or concrete. This is why a signal from an AM radio station can be received on the other side of an ocean, whereas the FM radio channels we listen to in a car must be changed every so often as we move along the road.

## 10.2 WIRELESS NETWORK ARCHITECTURES

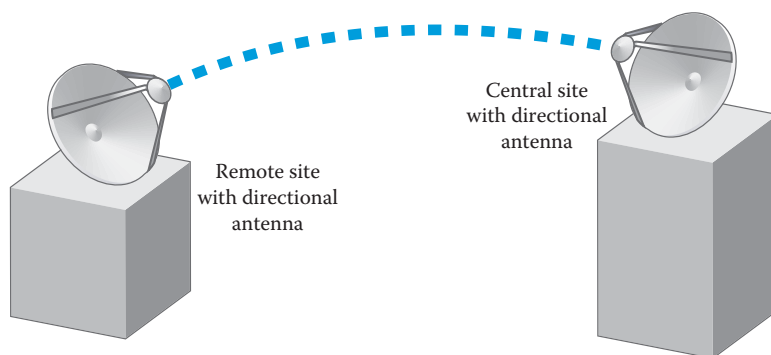
Depending on the application, wireless networks can be designed in different ways. In some scenarios, such as for a single camera in a parking lot, a point-to-point wireless connection would be the best choice for relaying video to the building where the video surveillance system is located. In a surveillance application with hundreds of cameras, a point-to-multipoint or even a mesh wireless network may be more appropriate. These three types of wireless architectures are explained in more detail in the following sections.

### 10.2.1 Point-to-point network

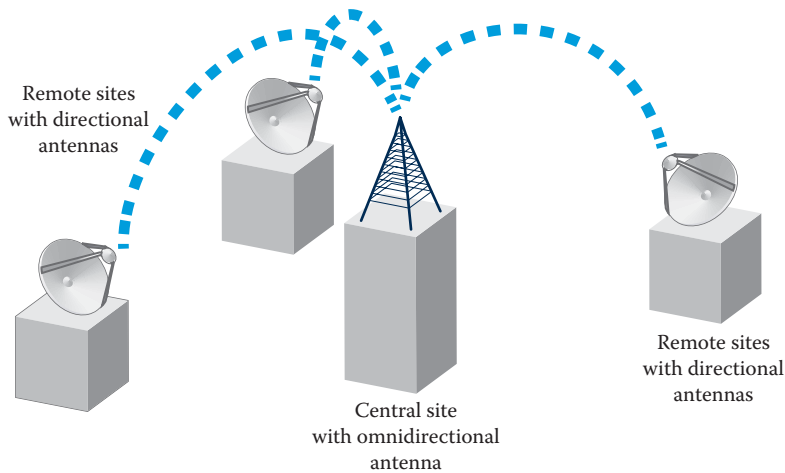
A point-to-point network (sometimes abbreviated as PTP or P2P) is the simplest wireless network, as it transmits information from one point to another (Figure 10.5). Because of the directional nature of the system, directional antennas are used to provide the highest bandwidth for the link. The system can also be adjusted for minimal interference and the highest level of security. If there is direct line of sight, it is also possible to use a high-performance optical link, for example, between two buildings located on opposite sides of a highway.

### 10.2.2 Point-to-multipoint network

A point-to-multipoint network (sometimes abbreviated PTMP or P2MP) is the most common type of wireless network (Figure 10.6). One example of such a network is an FM radio station transmitting radio signals to many receivers. Because of the nature of the network, the central point uses an omnidirectional antenna. The surrounding points use directional antennas, unless they are mobile (e.g., in a car), in which case an omnidirectional antenna is preferred. A point-to-multipoint network can be of broadcast and simplex nature, as in the case of FM radio, or it can be full duplex, with data being sent in both directions, as in a regular WLAN application.



**Figure 10.5** A point-to-point wireless network.

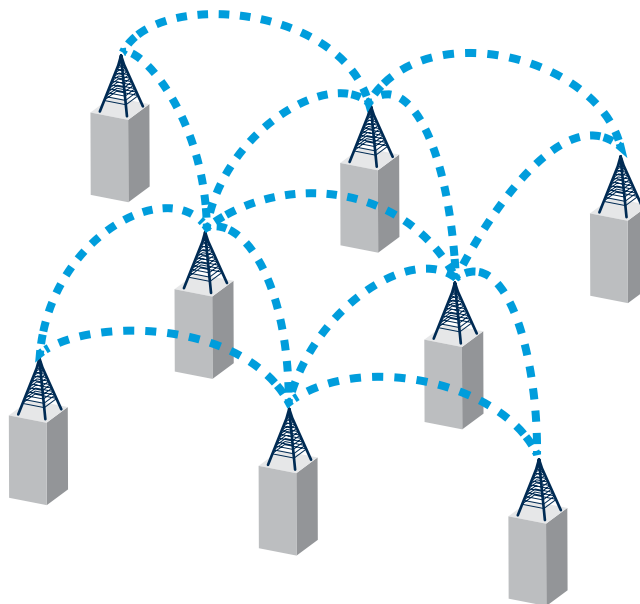


**Figure 10.6** A point-to-multipoint wireless network.

### 10.2.3 Mesh network

A wireless mesh network (Figure 10.7) is characterized by several connection nodes that provide individual and redundant connection paths from one to another. To accommodate this, special routing protocols are used to guarantee data exchange through the most appropriate connection path. When selecting a path, factors such as bandwidth, transfer errors, and latency are taken into account. The number of nodes between two data points is defined as the number of *hops*. The more hops, the longer the latency. Keeping latency and the number of hops to a minimum is important in applications such as live video and particularly in cases where PTZ cameras are used.

A wireless mesh network should be capable of managing itself. This means that if a node goes down, the system will automatically set up a new path between two points. Wireless mesh networks have



**Figure 10.7** A wireless mesh network.

now developed to the point where they now are a viable solution for video surveillance applications. Most current solutions are based on standard 802.11 technologies, with proprietary routing and security protocols.

## 10.3 802.11 WLAN STANDARDS

The most common wireless standard for data networks is 802.11, which was published in 1997 by the IEEE and specifies a media access control layer and different physical layers. The original two transfer protocols for wireless communication described by the 802.11 standard are the frequency hopping spread spectrum and the direct sequence spread spectrum (DSSS), both of which provide data transfer at 1 and 2 Mbit/s. Both protocols use the 2.4 GHz frequency spectrum, which is license-free all over the world, although in some regions the signal strength is limited to 20 dBm (100 mW). The 802.11 standard has also been improved and expanded, with the most important extensions outlined in the following sections.

### 10.3.1 802.11b extension

The 802.11b extension was approved in 1999. It uses DSSS in the 2.4 GHz range, giving data rates up to 11 Mbit/s. Until 2004, most WLAN products were based on 802.11b, and products supported data rates of 1, 2, 5.5, and 11 Mbit/s, depending on the distance.

### 10.3.2 802.11a extension

The 802.11a extension was also approved in 1999 and is based on the orthogonal frequency division multiplexing (OFDM) protocol. It operates in the 5 GHz frequency range and allows data rates of 6, 9, 12, 18, 24, 36, and 54 Mbit/s. Although higher data rates are possible, 802.11a also has some drawbacks and never became very popular.

### 10.3.3 802.11g extension

The 802.11g extension was approved in 2003 and quickly established itself as the new standard, replacing 802.11b. Wireless products that are 802.11b/g compliant use the 2.4 GHz frequency and provide data rates up to 11 Mbit/s with DSSS and up to 54 Mbit/s with OFDM.

### 10.3.4 802.11n extension

The 802.11n extension was approved in 2009 and enables data rates of up to 600 Mbit/s. These high rates are achieved with the multiple-input, multiple-output protocol, where multiple antennas and spread paths for the electromagnetic waves are used to provide several transfer routes that can be used in parallel to make the high rates possible. An 802.11n wireless network can operate in both the 2.4 and the 5 GHz band but will provide the highest capacity when used exclusively at 5 GHz, thanks to the many nonoverlapping channels and lesser degree of interference in this band.

### 10.3.5 802.11ac extension

The IEEE 802.11ac extension was approved in January 2014, building on and improving the performance of the 802.11n standard, although this later standard operates exclusively on the 5 GHz band. Depending on operating conditions, data throughput can be 1.7–2.5 Gbit/s. The IEEE 802.11ac extension also includes beamforming, which can direct radio signals at specific devices, increasing overall throughput and reducing power consumption. Products referred to as 802.11b/g/n/ac are compatible with all four extensions.

### 10.3.6 802.11s extension

The 802.11s extension, approved in 2011, makes it possible to create vendor-neutral mesh networks, with interoperability between many different types of devices from different manufacturers.



## 10.4 BASICS OF 802.11 NETWORKS

Different types of networks can be set up using the 802.11 technology. The following sections describe the basic 802.11 topologies and the frequencies in which they operate, as well as the significance of channels.

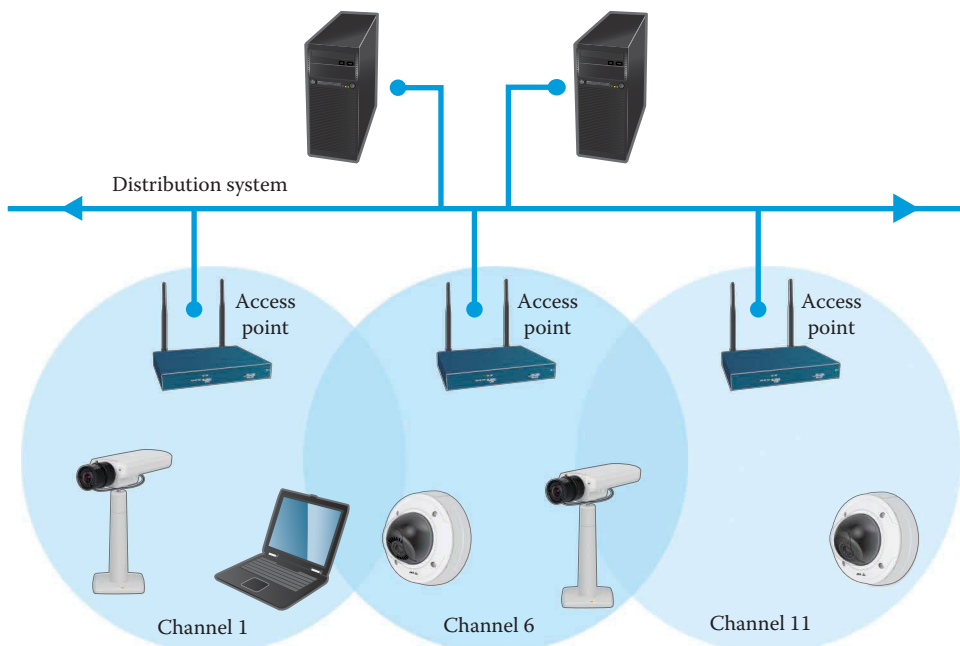
### 10.4.1 Infrastructure network

An access point has an Ethernet interface and at least one WLAN interface. Multiple access points can be used to cover a large area, whereby the individual access points must be positioned so that neighboring wireless cells slightly overlap one another to guarantee seamless transmission, as shown in Figure 10.8. In a case like this, it is important that the wireless cells in the immediate vicinity use different channels to avoid interference. After fulfilling these requirements, a user with a wireless device can move around the WLAN's complete coverage area without losing the connection to the network or suffering performance losses.

Moving between wireless cells is known as roaming. The access points that form a network are identified by the service set identifier (SSID). The SSID must be configured at each access point, and the nodes detect which access points belong to which network using the SSID, that is, which access points a node can associate with while roaming. If a device such as a network camera is to be integrated into an infrastructure network, both the infrastructure operating mode and the SSID should be selected in order to connect to the right WLAN network.

### 10.4.2 802.11 frequencies

As discussed previously, an 802.11 WLAN operates in the 2.4 or 5 GHz range. Both frequencies have their advantages and disadvantages. Networks that operate in the 2.4 GHz frequency are still the most common, but the greater throughput of 802.11n and 802.11ac also makes the 5 GHz band a popular choice. One of the major disadvantages of the 2.4 GHz frequency is that many other wireless technologies, such as Bluetooth, also use this frequency, with interference and reduced data rates as a result.



**Figure 10.8** To provide good coverage, the areas of the access points should overlap slightly.

Although the 5 GHz frequency range offers less interference, there are also disadvantages. The higher the signal's frequency, the lower the range. This is because the damping of electromagnetic signals increases as a function of the increased frequency. This can be particularly noticeable when signals must penetrate barriers such as walls or furniture. The signal range on the latest 5 GHz access points is much improved compared to earlier models, which required more units for the same area than comparable models using the 2.4 GHz range.

### 10.4.3 Channels

The 2.4 GHz range is divided into 14 channels. Channels 1–11 may be used freely everywhere, whereas the use of channels 12, 13, and 14 is restricted or may only occur at low power.

Channels are divided in such a way that neighboring channels overlap. When setting up a WLAN in the 2.4 GHz range, the open channels should be separated by several clear channels to avoid interference. This means that only three independent wireless channels will be open in a single wireless cell, for example, channels 1, 6, and 11 and that these separated channels must be shared between the access points.

## 10.5 WLAN SECURITY

---

When data are transferred over a wireless link, it is important to take into account that the fixed boundary available in a wired LAN does not exist in the wireless alternative. In theory, anyone within range of a WLAN could attack the network and intercept data. Consequently, security becomes even more important in preventing unauthorized access to data and the network. This section discusses some of the most common security technologies in wireless networking.

### 10.5.1 Wired Equivalent Privacy

The basic 802.11 standard includes Wired Equivalent Privacy (WEP), a security protocol intended to provide a level of security comparable to that available for data transferred by cable. WEP is used to encrypt and authorize data. Encoding is based on a symmetrical encryption protocol, which is based on RC4 Stream Encryption from RSA. All nodes and access points that exchange data must have the same secret WEP key so that the data can be encrypted and decrypted. Depending on the WEP version, various key lengths are used. In the case of WEP40, the key length is 40 bits, and in the case of WEP128, the key length is 128 bits.

Unfortunately, security loopholes in this protocol mean that WEP cannot be considered an adequate security method and its use cannot be recommended.

### 10.5.2 Temporal Key Integrity Protocol

The Temporal Key Integrity Protocol (TKIP) was introduced to provide security reliably through firmware or driver updates for previously installed wireless products. TKIP addresses the security loopholes in WEP and is software implemented on top of WEP.

### 10.5.3 Advanced Encryption Standard

The Advanced Encryption Standard (AES) is an encryption algorithm recommended by the National Institute of Standards and Technology. This algorithm provides a high level of security and efficiency and is implemented in hardware, that is, the router chip. AES in WLAN applications uses a key length of 128 bits, although it also supports 192 and 256 bit keys.

### 10.5.4 Preshared key

TKIP and AES both require keys from which a session key is derived. The preshared key (PSK) protocol or authentication to 802.1X can be used to derive a session key. PSK is simpler than authentication to 802.1X. When using PSK, the key derives from a passphrase. The passphrase has a length

The screenshot shows a web-based configuration interface for wireless networks. On the left is a sidebar menu with options: Basic Setup, Video, Live View Config, Detectors, Applications, Events, Recordings, Languages, System Options, and About. The 'Basic Setup' section is expanded, showing sub-items: Instructions, 1 Users, 2 Wireless (selected), 3 TCP/IP, 4 Date & Time, and 5 Video Stream.

The main content area is titled 'Wireless' and contains two sections:

**Status of Wireless Networks**

SSID	Mode	Security	Channel	Signal strength	Bit rate
axis-acc-env3	Master	WPA2-PSK	1	65 %	
axis-acc-env4	Master	WPA2-PSK	1	65 %	
axis-acc-env5	Master	WPA2-PSK	1	60 %	
axis-avhs-env1	Master	WPA2-PSK	1	65 %	
axis-avhs-env2	Master	WPA2-PSK	1	65 %	
axis-mobapp-env1	Master	WPA2-PSK	1	70 %	

Below the table is a 'Refresh' button.

**Wireless Settings**

☐ Enable congestion control  
☒ Enable WLAN pairing button

SSID:

Security:

Network type: ☒ Master

Passphrase:

Warning! Passphrases and keys saved here will be sent to the AXIS M1004-W in plain text.

**Figure 10.9** Depending on the type of security used, various parameters must be configured.

of 8–63 characters and must be entered manually into each node and access point. A PSK passphrase is secure when based on a meaningless string of letters, numbers, and special characters, for example, 3aRs5%3?&d48fgH67, so that a so-called dictionary attack cannot take place.

### 10.5.5 802.1X

Authentication to 802.1X can be used as an alternative to PSK. In 802.1X, the individual nodes must identify themselves to a Remote Authentication Dial-In User Service (RADIUS) server before they can transfer data over a network. If authentication is successful, a *pairwise master key* is generated, from which the session key is derived. The use of 802.1X requires a centralized RADIUS server, which has the advantage of being simple to manage. See Section 11.8.3 for more on 802.1X.

### 10.5.6 WiFi Protected Access®

When the security risks of WEP became apparent, the Wi-Fi Alliance® was forced to react. In 2003, the alliance defined Wi-Fi Protected Access® (WPA®) as the new security standard, using TKIP as the encryption technology. WPA2 began to be used in 2004 and corresponds to the AES implementation of 802.11i. Despite claims of successful WPA2 cracking, the standard is still considered safe for most applications, as long as settings and passphrases are sufficiently strong (Figure 10.9).

## 10.6 OTHER WIRELESS SOLUTIONS

In addition to the 802.11 WLAN standards, there are many other technologies available for the wireless transfer of data. Some of these are described in the following sections.

### 10.6.1 Bluetooth®

Bluetooth was originally a wireless solution that enabled a data transfer rate of 1 Mbit/s in the 2.4 GHz frequency range over short distances. The original key application for Bluetooth was to link peripherals wirelessly over a distance of up to 10 m (33 ft). The standard has been continually extended so that today data rates can reach 24 Mbit/s and distances of up to 100 m can be bridged. Bluetooth is not commonly used in video surveillance applications.

## 10.6.2 Universal Mobile Telecommunications System

The Universal Mobile Telecommunications System (UMTS) is a mobile communications standard of the third generation and includes a multitude of technologies. The first technology generation to offer bitrates high enough for data transfer was 3G and its use is now widespread.

The 4G standard is specified by the International Mobile Telecommunications Advanced specification. This requires a 4G technology to provide peak bitrates of up to 100 (Mbit/s) for mobile applications (e.g., in moving vehicles) and up to 1 Gbit/s for more-or-less stationary use. In its various forms, 4G is still being implemented and includes technologies such as Worldwide Interoperability for Microwave Access (WiMAX) (see Section 10.6.3). The future of 4G is known as LTE (Long Term Evolution), which is not so much a technology as a path for the development goals for this standard. The data rates available in these wireless networks make them of great interest to video surveillance applications.

## 10.6.3 Wireless interoperability for microwave access

WiMAX is a long-distance mobile communications technology specified in the IEEE-802.16 standard. The target market is metropolitan area network applications, where WiMAX is expected to compete with UMTS. The achievable data rate will depend on the distance. It is expected that many mobile communications providers in the future will offer WiMAX.

## 10.7 PERFORMANCE OF WIRELESS NETWORKS

---

The process of transferring data wirelessly is much more difficult than transferring it by cable, and there are significant data overheads for management, error recognition, and security. In the case of wired Ethernet, the overhead is relatively small; that is, the net data rate is close to the gross data rate.

The net data rate for a WLAN often amounts to considerably less than the gross data rate and is sometimes as low as 50% of the specified gross. When designing a wireless network for a video surveillance application and calculating the required bandwidth, the performance of the network must be taken into consideration to ensure that the desired frame rate is achievable.

## 10.8 BEST PRACTICES

---

For applications where a fixed data connection is not available and is too costly to install, a wireless network may be the only solution. Before selecting the technology and implementing a wireless network, there are some important things to consider:

- *How many nodes will be networked?* For two nodes only, a point-to-point connection may be appropriate, and several proprietary and standard solutions are available. If multiple nodes are involved, a multipoint or mesh network solution should be used.
- *What is the distance to be bridged?* The higher the frequency, the shorter the distance that can be covered. For very long distances, solutions in the 900 MHz range may be appropriate.
- *Are there any obstacles in the area of the wireless network?* A standard wireless network may have a range of 30–100 m indoors, but up to 300 m outdoors. Buildings, trees, and other obstacles will damp the wireless signal and limit its reach. The fewer the obstacles in the signal's path, the greater the distance that can be bridged.
- *How much bandwidth is required?* Most wireless solutions specify the maximum bandwidth. The real bandwidth depends on the distance and obstacles and may be much less than the specified maximum.
- *What is the risk of interference?* It is important to determine whether other wireless networks will be installed in the same area. Check existing equipment in the area with a spectrometer, which will be able to indicate the frequencies in use.

- *How secure does the connection need to be?* Wireless means sending data over the open air. If used for video surveillance, it is important to make the data secure by using the appropriate security protocols, such as 802.1X. Remember that adding security protocols normally decreases performance.

Although a wireless network can replace a wired network, there is still a need to supply power to all nodes. Power over Ethernet (PoE) removes the need for a separate power cable but retains the data cable. For most indoor video surveillance applications, PoE is more cost effective and more secure than wireless. For more information on PoE, see Section 9.6.



# Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>



# CHAPTER 11

## Networking technologies

Networking technology has experienced tremendous development over the past decades. Today, billions of people globally are using networking technology, tens of billions of devices are connected to networks and the internet, and the numbers are growing fast. The technology is quite complex, but most users are not exposed to this. When a user switches on their laptop, a series of networking technologies is automatically initiated to ensure that the laptop gets an Internet Protocol (IP) address and that the network communication is secure. The IP is the common denominator in network technology, and although it was originally designed for military communications, today it is also used in small home networks, enterprise local area networks (LANs), and the internet for applications such as email, web browsing, telephony, and network video.

Many different protocols come into play when data are transferred securely from one networked device to another. The best way to understand how the different protocols interact is to examine the Open Systems Interconnection (OSI) communication model, as explained in the first section of this chapter. The IP is discussed extensively in the middle of the chapter, followed by a section on a topic that is becoming increasingly important as network use continues to grow, network security—sometimes referred to as cybersecurity.

### 11.1 OSI REFERENCE MODEL

---

Data communication between open systems is described using the OSI Reference Model, which is composed of seven layers (Figure 11.1). Each layer provides specific services and makes the results available to the next layer. To provide a service, each layer utilizes the services of the layer immediately below it. Communication between layers occurs via specific interfaces. Each layer must follow certain rules, known as protocols, to perform its services. It is important to note that the OSI is not a protocol in itself, but rather a model used to understand the function of the included protocols.

Communication between two systems always occurs on the same layer. This is known as virtual communication, and it can only occur when the same protocols are implemented in the corresponding layers of each system. A system passes data downward to the lowest vertical layer (the physical) and then transfers it horizontally to the other system, where the data are then passed upward. In this way, the data reach the corresponding layer on the other system and communication takes place.

Layer 7—Application	Data
Layer 6—Presentation	Data
Layer 5—Session	Data
Layer 4—Transport	Segments
Layer 3—Network	Packets
Layer 2—Data Link	Frames
Layer 1—Physical	Bits

**Figure 11.1** The Open Systems Interconnection Reference Model consists of seven layers. Each layer performs a particular service.

**11.1.1 Layer 1: The physical layer**

The physical layer is the lowest layer and it provides services that support the transmission of data as a bitstream over a medium, such as a wired or wireless transmission link. This layer describes the transmission medium and its physical characteristics, as well as the mechanical and electrical means that permit a physical connection to the transmission medium. This layer also defines the form for electrical signals, optical signals, or electromagnetic waves, as well as the connectors and sockets needed for a network cable connection.

**11.1.2 Layer 2: The data-link layer**

The data-link layer provides data transmission and controls access to the transmission medium, by combining data into units known as frames. These frames are provided with a checksum, which the recipient uses to detect possible transmission errors. According to the Institute of Electrical and Electronics Engineers (IEEE), the second layer is divided into two sublayers, with the upper range corresponding to the Logical Link Control (LLC) and the lower part corresponding to the media access control (MAC). The LLC is used in the same way by all IEEE network technologies and simplifies data exchange. The MAC controls access to the transmission medium and depends on the network technology used.

Examples of protocols and standards are IEEE 802.2 (LLC), IEEE 802.3 (Ethernet MAC), and 802.11 (WLAN MAC).

**11.1.3 Layer 3: The network layer**

The network layer performs the actual data transfer, by routing and forwarding data packets between systems. The most important tasks of this layer include the creation and administration of routing tables, and it provides options for communicating beyond network boundaries. The data in this layer are assigned destination and source addresses, which are used as the basis for targeted routing. Addressing in the third layer is independent of addressing at lower levels, which means that routing can extend to multiple, logically structured networks.

Examples of protocols that operate in this layer are the IP (IP version 4 [IPv4] and IP version 6 [IPv6]), Routing Information Protocol, and Internet Protocol Security (IPSec).

**11.1.4 Layer 4: The transport layer**

The function of the transport layer is to provide reliable data transfer service to Layer 5 and above. The transport layer controls the reliability of a wired or wireless link through flow control and

error control. Some protocols that perform in this layer are oriented toward state and connection, meaning that they can keep track of segments and retransmit those that fail.

Examples of protocols in this layer are the Transmission Control Protocol (TCP) and the User Datagram Protocol (UDP).

### 11.1.5 Layer 5: The session layer

The session layer provides an application-oriented service and takes care of the process communication between two systems. Process communication begins with the establishment of a session, which provides the basis for a virtual connection between two systems.

Examples of protocols that provide Layer 5 functions include Remote Procedure Call and Network File System.

### 11.1.6 Layer 6: The presentation layer

The presentation layer converts system-dependent data formats, such as ASCII, into an independent format, thus permitting syntactically correct data exchange between different systems. The tasks of this layer also include data compression and encryption. The presentation layer ensures that data sent by the application layer of a system can be read by the application layer of another system.

Examples of protocols that provide Layer 6 functions include Telnet and Apple Filing Protocol.

### 11.1.7 Layer 7: The application layer

The application layer is the highest layer of the OSI model. It makes functions such as web, file, and email transfer available to applications. The actual applications, such as web browsers or Microsoft® Outlook®, exist above this layer and are not covered by the OSI model.

Typical protocols in this layer are the File Transfer Protocol (FTP), Simple Mail Transfer Protocol (SMTP), and Hypertext Transfer Protocol (HTTP).

## 11.2 TCP/IP REFERENCE MODEL

---

The TCP/IP reference model can be used to understand protocols and how communication takes place. TCP/IP stands for Transmission Control Protocol/Internet Protocol, and in the TCP/IP reference model, the protocols fall into four different layers, which correspond to the seven layers in the OSI model as shown in Figure 11.2.

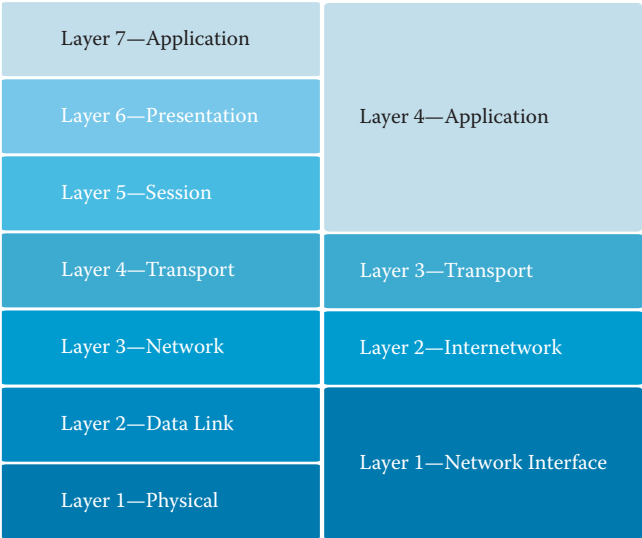
### 11.2.1 Internet Protocol

The IP is a Layer 3 protocol in the OSI model and a Layer 2 protocol in the TCP/IP Reference Model. It is the key protocol in most of today's networked applications.

For a server to operate on the internet, it must have its own, individual, public IP address. The Internet Corporation for Assigned Names and Numbers is a nonprofit global organization that has been designated to conduct the registrar accreditation process. An IP address can be assigned by an internet service provider (ISP), which can allocate either a dynamic IP address, which can change from session to session, or a static address, which is normally provided for a monthly fee.

IP offers a connectionless service that splits data into IP datagrams before they are transmitted. *Connectionless* means that IP does not guarantee whether or in what sequence IP datagrams will arrive at a recipient. Because IP operates on a connectionless basis, individual IP datagrams can be routed to a recipient via different paths. Guaranteed data transmission and reassembly into the correct sequence are performed by protocols in the transport layer of the OSI and TCP/IP models.

There are currently two versions of IP: IPv4 and IPv6. IPv6 is slowly replacing IPv4, but this older version still provides the vast majority of the service today. The significant difference between the two IP versions is the size of the address space, which is substantially larger in IPv6.



**Figure 11.2** The TCP/IP reference model can be split into four layers (right), which is comparable to the seven layers in the Open Systems Interconnection model (left).

### 11.2.2 IPv4 addresses

An IPv4 address is 32 bits long, meaning that  $2^{32}$ , or 4.3 billion (4,294,967,296), unique IP addresses can be assigned. To make them more readable, IP addresses are grouped into four blocks of 1 byte (8 bits) each. The individual blocks (referred to as octets) are separated by a dot, and each block represents a decimal value between 0 and 255 (e.g., 85.235.16.37). This is known as dot-decimal notation.

Each IP address consists of a network ID and a host ID. The network ID represents the logical IP network in which the host resides. The host ID represents the host itself. All devices with the same network ID will reside on the same network segment, meaning that these hosts can communicate directly with each other without the use of a router to forward the traffic to the correct network.

IP addresses can be classified in five groups, referred to as classes (Table 11.1). These classes are based on the number range of the first octet of the IP address. Each class defines how many network and host addresses it has available. IP addresses used in LANs and WANs come from classes A, B, and C, with class C being used in most applications. Classes D and E are for multicast and experimental uses.

Note, however, that the class-based addressing model of IPv4 is not commonly used today. Instead, Classless Interdomain Routing (CIDR) is used, which permits a more efficient use of the limited supply of IPv4 addresses.

**Table 11.1** Classes of Internet Protocol addresses

Class	Value range of first byte	Bytes for net ID	Number of networks	Bytes for host ID	Number of hosts
A	1–126	1	126	3	16,777,214
B	128–191	2	16,384	2	65,534
C	192–223	3	2,097,152	1	254
D	224–239	Multicast addresses	N/A	N/A	N/A
E	240–254	Reserved	N/A	N/A	N/A

11.2.3 Subnets

An IPv4 address is divided into two parts: the network ID and the host ID. The subnet mask determines the blocks of an IPv4 address that define the network and host identifiers. A subnet mask has a length of 32 bits and is also represented by dot-decimal notation (e.g., 255.255.255.0).

When determining the network and host IDs within an IP address, the computer converts the IP address and subnet mask into binary. Table 11.2 shows the binary-to-decimal numbering conversions.

When viewed as a binary number, the bits of the subnet mask are a contiguous group of 1’s followed by a contiguous group of 0’s. The bits of an IP address that correspond to the group of ones in the subnet mask represent the network ID. The bits of an IP address that correspond to the group of zeros in the subnet mask represent the host ID. Only subnet masks with contiguous bits are allowed.

For example, a permitted subnet mask would be

255.255.192.00 (11111111.11111111.11000000.00000000)  
255.255.255.0 (11111111.11111111.11111111.00000000)

The following subnet mask contains noncontiguous bits and would not be permitted:

255.255.230.0 (11111111.11111111.11100110.00000000)

Table 11.3 provides some examples of IP addresses that are broken down into the network and host ID. In Example 1, the subnet mask indicates that the first three blocks of digits in the IP address define the network ID, and the last block, the host ID. In Example 2, the subnet mask indicates that the first two blocks define the network ID, whereas the last two blocks define the host ID.

If a host wants to transmit data, it determines the network ID of the destination IP address by means of the subnet mask. If the destination IP address has the same network ID as the host, then the host sends the data directly to the destination.

If the network ID of the destination IP address is located on another subnet, the host instead sends the data to the default gateway as defined in the host. The default gateway is the IP address of a router that forwards data to the correct network. There is an alternative notation based on IPv4-CIDR

Table 11.2 Permitted values for subnet mask bits

128	64	32	16	8	4	2	1	—
1	0	0	0	0	0	0	0	128
1	1	0	0	0	0	0	0	192
1	1	1	0	0	0	0	0	224
1	1	1	1	0	0	0	0	240
1	1	1	1	1	0	0	0	248
1	1	1	1	1	1	0	0	252
1	1	1	1	1	1	1	0	254
1	1	1	1	1	1	1	1	255

Table 11.3 Examples of Internet Protocol addresses broken down into network ID and host ID

	Example 1	Example 2
IP address	192.168.1.5	10.50.88.129
Subnet mask	255.255.255.0	255.255.0.0
Network ID	192.168.1.0	10.50.0.0
Host ID	0.0.0.5	0.0.88.129

for representing the IP address and subnet mask, and it uses a suffix to indicate the length of the subnet mask (the number of 1's). The number of bits used for the subnet mask is appended to the IPv4 address as a decimal number followed by "/". For example, 192.168.12.23/24 corresponds to the IP address 192.168.12.23 with the subnet mask 255.255.255.0.

### 11.2.4 Network address translation

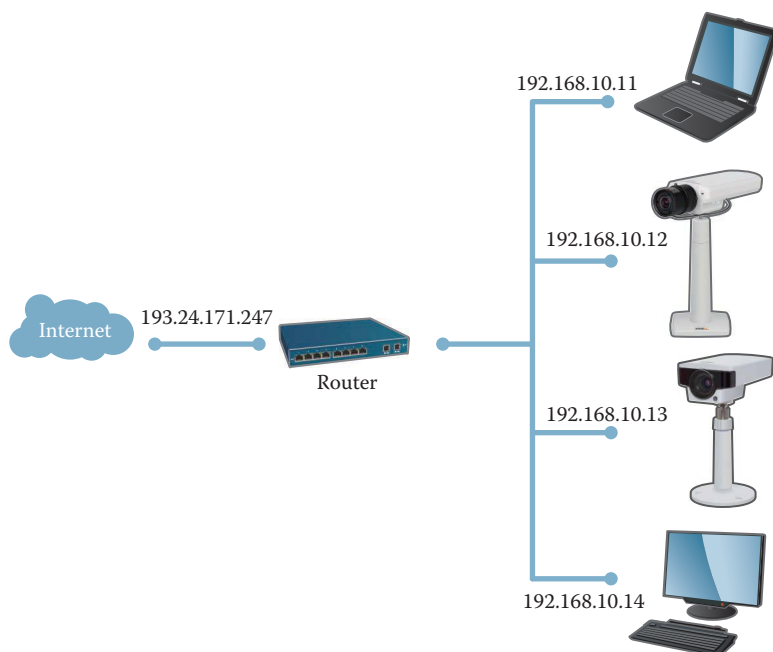
Three address ranges are defined for use exclusively for private purposes. These private IP addresses are 10.0.0.0 to 10.255.255.255, 172.16.0.0 to 172.31.255.255, and 192.168.0.0 to 192.168.255.255. These addresses can be used only on private networks and are not allowed to be forwarded through a router to the internet. If internet connectivity is required, this can be provided using a technique called Network Address Translation (NAT), which operates on Layers 3 and 4 of the OSI model.

In such a case, IP datagrams are forwarded by a NAT-enabled router that translates the private IP address into a public IP address. The public IP address is the one allocated by the internet provider. Larger private networks with multiple servers also can use the same readdressing technique. The router manages the connection between the private network and the internet (Figure 11.3).

### 11.2.5 Services and port numbers

Data sent from one system to another must be associated with a particular service or application so that the receiving server knows how to process it. This is done by associating the data with a service or process defined by, or mapped to, a particular port number on a server. For example, a service that runs on a web server is typically mapped to port 80 on a computer (Figure 11.4).

A port number is 16 bits long, that is, values from 0 to 65535. Certain applications use port numbers preassigned to them by the Internet Assigned Numbers Authority. These numbers are in the range 0–1023 and are called well-known registered ports. The ports 1024–49151 are known as registered ports. When requested to do so, manufacturers of applications can register ports from this range for their own protocols. The remaining ports, 49152–65535, are called private ports and can be used freely as they are not registered and not assigned to any particular application.



**Figure 11.3** Multiple servers can connect to the internet through a unique public internet Protocol address using a Network Address Translation-enabled router.



Basic Setup

Video & Audio

Live View Config

Detectors

Applications

Events

Recordings

Languages

System Options

Security

Date & Time

Network

TCP/IP

Basic

Advanced

SOCKS

QoS

SNMP

UPnP™

RTP

Bonjour

Storage

Ports & Devices

Maintenance

Support

Advanced

About

Advanced TCP/IP Settings

DNS Configuration

Obtain DNS server address via DHCP

View

Use the following DNS server address:

Domain name: (use ; to separate names)

Primary DNS server:

Secondary DNS server:

NTP Configuration

Obtain NTP server address via DHCP

View

Use the following NTP server address:

Network address: (host name or IP address)

Host Name Configuration

Obtain host name via IPv4 DHCP

View

Use the host name: axis-accc8e02009e

Enable dynamic DNS updates

Register DNS name: (Axisproduct.example.com)

TTL: 30

Link-Local IPv4 Address

Auto-Configure Link-Local Address

View

HTTP

HTTP port: 80

HTTPS

HTTPS port: 443

NAT traversal (port mapping) for IPv4

NAT traversal is disabled.

Enable

Use manually selected NAT router: (LAN IP address)

Alternative HTTP port: 0 \*

\* If set to blank or 0, a port number will be set automatically upon enable.

FTP

Enable FTP server

RTSP

Enable RTSP server \*

RTSP port: 554

\* H.264 video streams will be unavailable if this is disabled.

Save

Reset

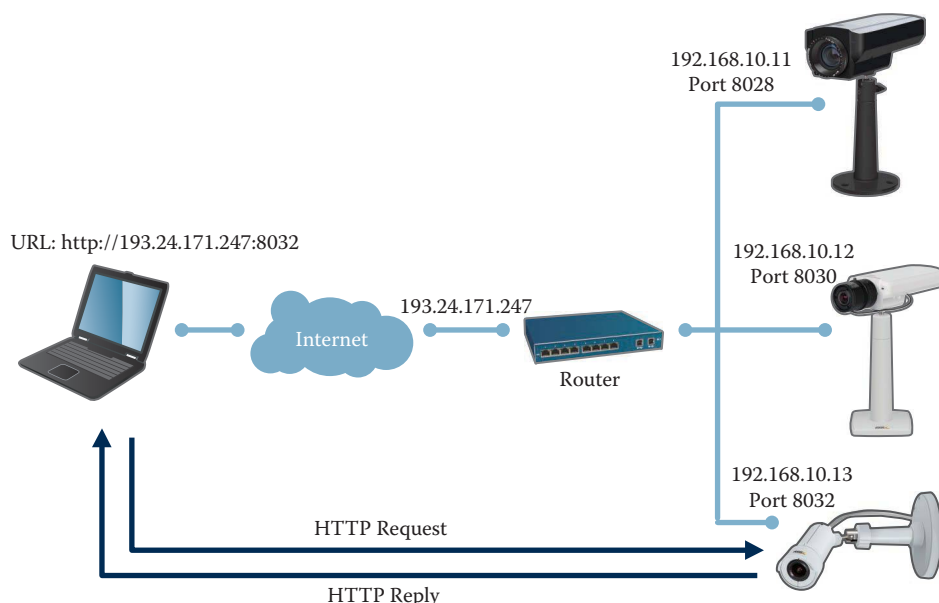
**Figure 11.4** In many network cameras, the port number can be set as required. Normally, the port is set to port 80 if the network camera is accessed as a web service over Hypertext Transfer Protocol.

Many network video products permit reconfiguration of the port numbers of individual services. For example, the port number of the web server service on network cameras can be changed from port 80 to a private port. Using a private port also means that it will be more complex to access a network camera through a web browser, which may actually be desirable in some applications.

11.2.6 Port forwarding

To configure access from the internet to cameras located on a private LAN, one possible technique is port forwarding, which is available in most routers today.

In port forwarding (see Figure 11.5), incoming data packets reach the NAT-enabled router by way of a public IP address. The router is configured to forward any data arriving at a predefined port to a specific host on the private network side of the router. The router then replaces the address of the incoming packet with a private IP address. To a receiving client, it looks like the source of the packets originated from the router. The reverse happens with outgoing data packets. The router replaces the private IP address of the source device with a public IP address before sending the data over the internet.



**Figure 11.5** Thanks to port forwarding, network cameras on a local network can be addressed individually over the internet.

To access a network camera over the internet, the public IP address of the router should be used together with the corresponding port number of the device on the private network. In the system illustrated in Figure 11.5, typing the URL `http://212.134.245.45:8032` into a browser would give the user access to a device with a private IP address of 192.168.10.13 Port 8032.

## 11.2.7 IPv6

So far, the shortage of IPv4 addresses has been managed by using techniques such as NAT. Now however, these addresses are no longer enough. Fortunately, IPv6 offers a solution to the problem: a huge number of IP addresses.

Other major advantages of IPv6 include IP autoconfiguration based on the MAC address, renumbering to simplify switching entire corporate networks between providers, faster routing, point-to-point encryption according to IPSec, and connectivity using the same IP address in changing networks (Mobile IPv6).

### 11.2.7.1 IPv6 addresses

An IPv6 address has a length of 128 bits, which means that it offers  $2^{128}$ , or approximately 340.28 sextillion ( $3.4 \times 10^{38}$ ) addresses. An IPv6 address is written in hexadecimal notation with colons dividing the address into eight blocks of 16 bits each. An example of an IPv6 address is

```
2001:0da8:65b4:05d3:1315:7c1f:0461:7847
```

To simplify how the address is represented, two consecutive colons can replace one or more 16-bit groups with the value 0000. However, the resulting address may contain two consecutive colons once only. These two addresses are equivalent to each other:

```
2002:0da8::1315:7c7a
```

```
2002:0da8:0000:0000:0000:0000:1315:7c7a
```

Leading zeroes in a 16-bit group also can be omitted; for example, `2002:0da8::0017:000c` can be written as `2002:da8::17:c`.

Address ranges in IPv6 are indicated by prefixes; subnets too are determined by the prefix. The prefix length (number of bits) is appended to the IPv6 address as a decimal number followed

by “/” (forward slash). Subnet masks as used in IPv4 do not exist in IPv6; instead, a notation similar to IPv4-CIDR is used.

In IPv6, the first 64 bits of the address are usually intended for network addressing, and the last 64 bits are for host addressing. For example, if a device has the following IPv6 address

```
2002:0da8:67f3:08a4:1511:aa56:0361:7a4f
```

then the device comes from the subnet

```
2002:0da8:67f3:08a4::/64
```

IPv6 enables a device to automatically configure its IP address using the MAC address (see Figure 11.6). In these cases, the prefix—the first 64 bits—is always the same; fe80 and the remaining 48 bits correspond to zeroes (fe80:0000:0000:0000). For the remaining 64 bits (suffix) of the IPv6 address, the MAC address of the system is converted into the Extended Unique Identifier-64 (EUI-64) numbering system. The result would be, for example,

```
fe80::1511:aa56:0361:7a4f
```

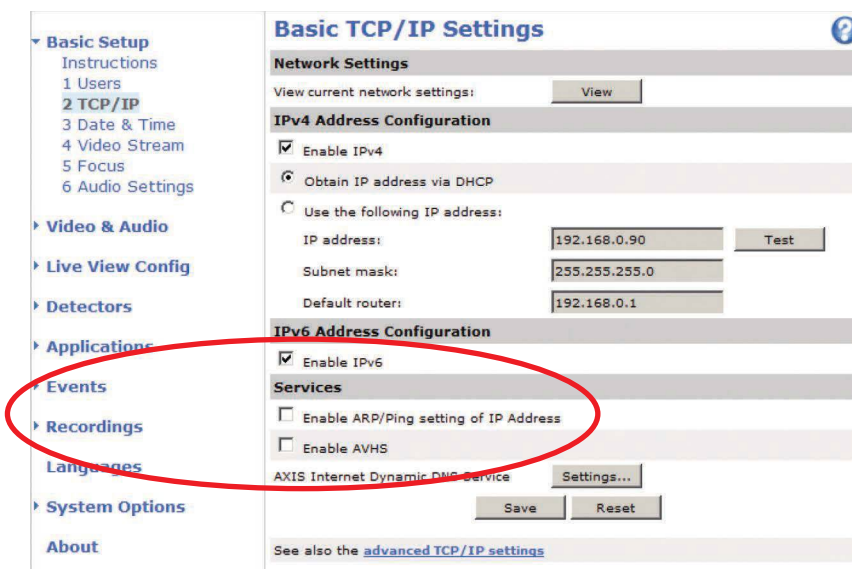
The MAC-based address permits a networked device to communicate on the local network. However, for communication over the internet, the first 64 bits of the IPv6 address must be adapted to the network address of the router, as allocated by the ISP. To do this, a device sends the router a corresponding host request and receives the necessary prefix of the public address block and additional information from the router. Using this information, the device can create the IPv6 address from the prefix and its suffix (EUI-64 address). Services such as Dynamic Host Configuration Protocol (DHCP) for IP address allocation and tasks such as the manual configuration of IP addresses are not required in IPv6.

The IPv6 address is enclosed in square brackets in a URL. An example of a correct URL is

```
http://[2002:0da8:67f3:08a4:1511:aa56:0361:7a4f]/
```

A specific port also can be addressed by changing the address as follows:

```
http://[2002:0da8:67f3:08a4:1511:aa56:0361:7a4f]:8081/
```



**Figure 11.6** Many network cameras have support for IPv6, which is becoming increasingly important in network video applications.

## 11.3 MANAGING IP ADDRESSES

In the early days of networking, most devices had a fixed IP address that was set and managed manually. As networks grew, so did the demand for techniques to manage networked devices and automate the tasks of setting and tracking IP addresses.

The following section describes the different ways of setting and managing IP addresses. IP addressing of data packets and the use of Domain Name System (DNS) are also explained.

### 11.3.1 Setting IP addresses

Any device on an IP-based network must have a unique and valid IP address. Setting the IP address can be done in two different ways:

- Manually using a static address
- Dynamically using DHCP

### 11.3.2 Manual address allocation

Setting IP addresses manually is labor intensive. Static addresses are normally used only in smaller system or for devices that need static address for particular reasons. One way to set a static IP address for a network camera is to access its built-in web pages (Figure 11.7).

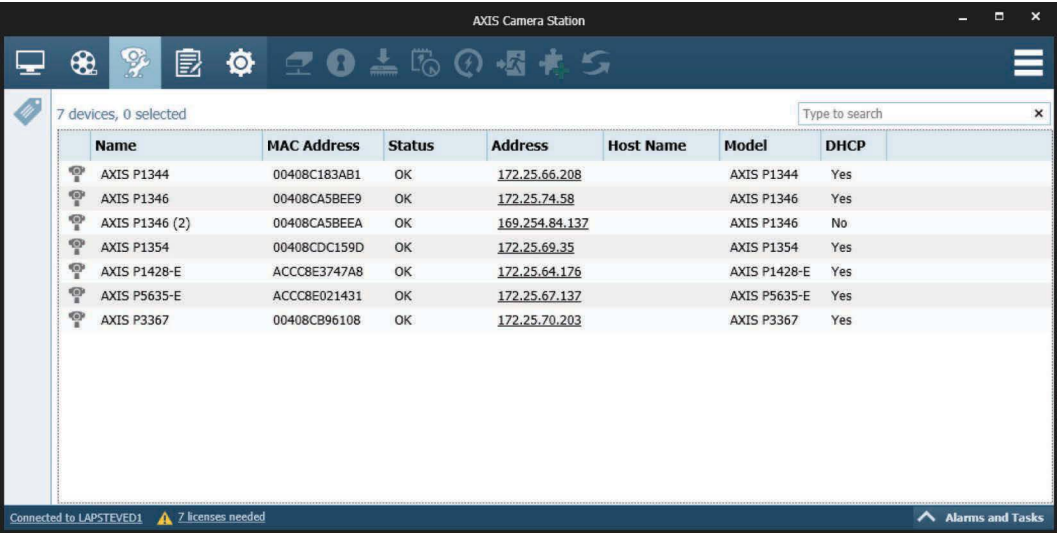
Many vendors offer tools not only for setting IP addresses but more importantly for finding and managing devices on a network. In a network video solution with potentially hundreds of network cameras, these tools are essential for effective management of the system (see Figure 11.8).

#### 11.3.2.1 Dynamic address allocation

The DHCP is the standard protocol used for the automatic assignment and management of IP addresses. A DHCP server manages a pool of IP addresses, which it can assign dynamically to a DHCP client upon request. The DHCP server can also provide other IP configuration parameters to DHCP clients, such as host names and DNS (see Section 11.3.4) server addresses. A DHCP server is the central point of client configuration management; it removes the need to maintain individual network configurations for each network client.

The screenshot displays the 'Basic TCP/IP Settings' web interface. On the left is a navigation menu with options like 'Basic Setup', 'Video & Audio', 'Live View Config', 'Detectors', 'Applications', 'Events', 'Recordings', 'Languages', 'System Options', and 'About'. The main content area is titled 'Basic TCP/IP Settings' and includes a 'Network Settings' section with a 'View' button. Below this is the 'IPv4 Address Configuration' section, which is circled in red. It contains a checked 'Enable IPv4' checkbox, a radio button for 'Obtain IP address via DHCP', and a selected radio button for 'Use the following IP address:'. Under this selection, there are input fields for 'IP address:' (192.168.0.90), 'Subnet mask:' (255.255.255.0), and 'Default router:' (192.168.0.1), with a 'Test' button next to the IP field. Below the IPv4 section is the 'IPv6 Address Configuration' section with an unchecked 'Enable IPv6' checkbox. The 'Services' section follows, with 'Enable ARP/Ping setting of IP Address' checked and 'Enable AVHS' unchecked. At the bottom, there is an 'AXIS Internet Dynamic DNS Service' section with a 'Settings...' button, and 'Save' and 'Reset' buttons. A link to 'advanced TCP/IP settings' is at the very bottom.

**Figure 11.7** Setting a static IP address from a camera's web interface.



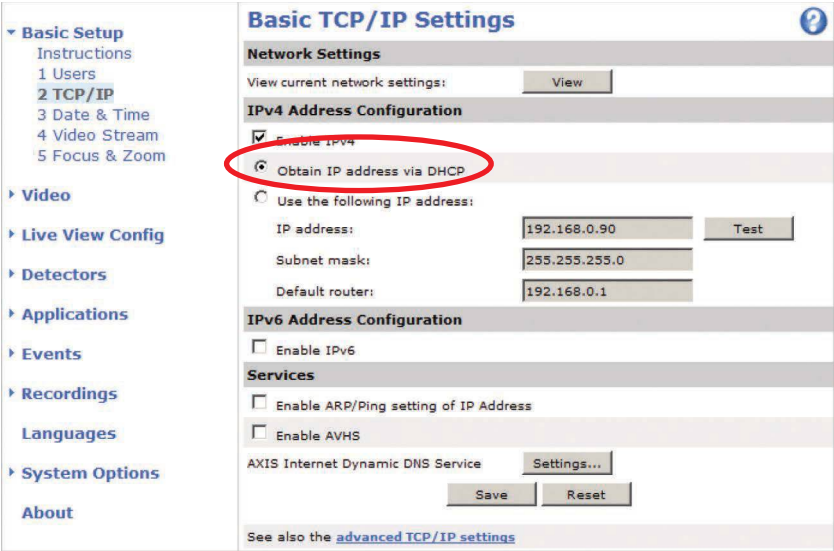
Name	MAC Address	Status	Address	Host Name	Model	DHCP
AXIS P1344	00408C183AB1	OK	172.25.66.208		AXIS P1344	Yes
AXIS P1346	00408CA5BEE9	OK	172.25.74.58		AXIS P1346	Yes
AXIS P1346 (2)	00408CA5BEEA	OK	169.254.84.137		AXIS P1346	No
AXIS P1354	00408CDC159D	OK	172.25.69.35		AXIS P1354	Yes
AXIS P1428-E	ACCC8E3747A8	OK	172.25.64.176		AXIS P1428-E	Yes
AXIS P5635-E	ACCC8E021431	OK	172.25.67.137		AXIS P5635-E	Yes
AXIS P3367	00408CB96108	OK	172.25.70.203		AXIS P3367	Yes

Figure 11.8 Most vendors provide tools for managing the IP addresses of networked devices.

When a DHCP client comes online, it sends a query requesting configuration from the DHCP server, which replies with an IP address, subnet mask, and other configuration parameters. In a typical configuration, the IP address provided by a DHCP server is leased to the client. Once half the lease time has expired, the client will request to renew the lease (see Figure 11.9).

11.3.3 Configuration-free networking

In many systems, especially small ones, setting and managing IP addresses is complex and cumbersome. To address these issues, several techniques have been developed, which simplify and automate IP addressing to the greatest possible extent. The two most well-known techniques are Universal Plug and Play (UPnP®) and Zeroconf (of which Apple’s Bonjour® is an implementation).



Basic Setup

Instructions

1 Users

2 TCP/IP

3 Date & Time

4 Video Stream

5 Focus & Zoom

Video

Live View Config

Detectors

Applications

Events

Recordings

Languages

System Options

About

Basic TCP/IP Settings

Network Settings

View current network settings: View

IPv4 Address Configuration

☒ Enable IPv4

☒ Obtain IP address via DHCP

☐ Use the following IP address:

IP address: 192.168.0.90

Subnet mask: 255.255.255.0

Default router: 192.168.0.1

Test

IPv6 Address Configuration

☐ Enable IPv6

Services

☐ Enable ARP/Ping setting of IP Address

☐ Enable AVHS

AXIS Internet Dynamic DNS Service Settings...

Save Reset

See also the advanced TCP/IP settings

Figure 11.9 Any networked device, such as a PC, must have a unique Internet Protocol address, which can be set dynamically by DHCP.

### 11.3.3.1 UPnP® and Zeroconf

Zeroconf is a component of UPnP. By using UPnP, Microsoft operating systems can automatically detect resources on a network, meaning that a network camera will automatically be listed under “Network” in the Windows® operating system.

With Zeroconf, networked devices attempt to independently allocate an IP address in the range from 169.254.1.0 to 169.254.254.255. If a system wishes to configure an IP address, it simply selects an address from the range using a random number generator based on system-specific information, such as the MAC address. After selecting an appropriate IP address, the system must first check to see if it is already in use by another device.

### 11.3.3.2 Bonjour®

Bonjour is another protocol for the announcement and discovery of services in an IP network. This protocol is based on open standards and was introduced by Apple. Using Bonjour, services can be discovered independently, and any necessary configuration can be performed automatically without user intervention. Bonjour is based on the exchange of multicast DNS packets sent via the UDP port 5353 (224.0.0.251).

In addition to multicast DNS, Bonjour is based on DNS Service Discovery (DNS-SD). DNS-SD is an expansion of the Domain Name Service as used for domain management. Bonjour is comparable to UPnP and Zeroconf.

Bonjour is appropriate for use in discovering network video products using Mac computers, but it can also be used as a discovery protocol for new devices in any network.

### 11.3.3.3 MAC and IP address resolution

IP addressing of data packets operates on the third OSI layer. Before data can be transported over a physical network, it must be packaged into frames, a process that occurs in the data-link (second) layer of the OSI model. Addressing on the second layer uses MAC addresses, which means that in addition to knowing the destination IP address, the sending host must also know the MAC address of the destination host—so that the destination IP address can be associated with a MAC address. To discover the MAC address of a given destination host, the sending host transmits a request using a protocol called Address Resolution Protocol (ARP).

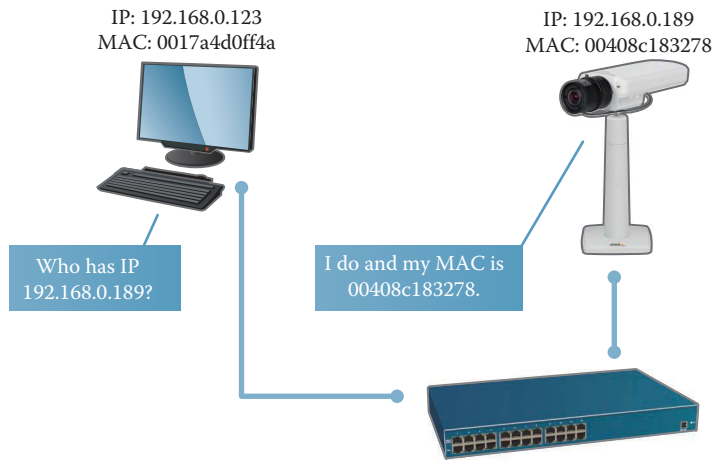
### 11.3.3.4 Address resolution protocol

The ARP is used to discover the MAC address of the destination host. An ARP request specifies the IP address of the destination host, in addition to the IP and MAC addresses of the sending host. When a switch receives an ARP request, it broadcasts it to all devices on the local segment. The devices then compare the destination address to their own IP addresses. The device with the corresponding IP address replies by disclosing its MAC address.

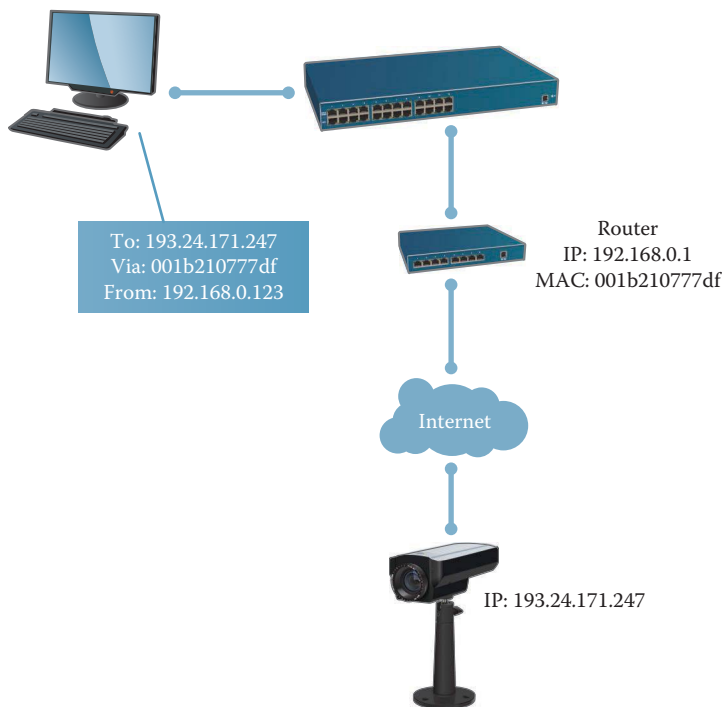
The requesting host then enters the MAC address it received, together with the associated IP address, into its ARP cache. The ARP cache is a table that temporarily stores the mapping of addresses, usually between MAC and IP addresses, in the RAM memory of a host. Before transmitting the data, the sending host checks whether or not it can resolve the necessary MAC address using its ARP cache. If this cannot be done, it sends out an ARP request (see Figure 11.10).

When a destination host resides outside a sending host’s own LAN, the sender must instead use the MAC address of its router as it is not able to discover the MAC address of the destination host (Figure 11.11). If the sending host does not know the MAC address of its router an ARP request is sent to discover it. Thus, to properly address a data packet that will be sent to another LAN, the sending host uses the MAC address of its router, together with the IP address of the destination host.





**Figure 11.10** To send data over a local network, Address Resolution Protocol is used so that the sending host can properly address the data packet with the destination host's IP and MAC addresses.



**Figure 11.11** To send data to a destination host outside a local network, the sending host uses its router's MAC address, together with the destination host's IP address.

### 11.3.4 Domain name system

The DNS converts domain names into their associated IP addresses, and it operates on the transport layer. This can be compared to a telephone book, which links the name of a person to a phone number. The DNS-to-IP address mapping can be done either manually in a host file or automatically, which is more common.

DNS caters to the human ability to better remember names than numbers. For example, the domain name *company.com* is much easier to remember than its associated IP address *193.24.171.247*.

DNS is used on a daily basis when a website is accessed by typing a domain name, such as `www.company.com`, into a web browser's address line. DNS then converts this name into the IP address, and communication between the systems takes place.

#### 11.3.4.1 Dynamic DNS

When a PC or network camera connects to the internet, the ISP will typically assign an unused IP address dynamically, through DHCP. This address can only be used for a short time, and several addresses may be used over the duration of a connection. When this happens, a Dynamic DNS (DynDNS) is used to keep track of a domain name's link to changing IPv4 addresses.

In DynDNS, the host record on the DNS server is updated whenever the host's IP address is changed. Depending on the host's capabilities and network configuration, either the host itself or the DHCP server sends this update.

Using DynDNS, a device's domain name (which never changes) can always be used to access the device, regardless of the IP address currently assigned to it.

## 11.4 DATA TRANSPORT

IP is the most important protocol in the network layer and the core protocol for any data communication. However, IP never acts alone but always together with protocols in the same or higher layers. In the transport layer (Layer 4 in the OSI model), the most common protocols are UDP and TCP. These protocols, along with the application protocols described in Section 11.5, are generally referred to as the IP or TCP/IP protocol suite.

### 11.4.1 User Datagram Protocol

UDP is a network protocol located on the fourth OSI layer; it provides a connectionless transmission service. The fact that the protocol is connectionless means that no connection is established between the sender and the receiver.

Messages are packaged into datagrams and transmitted. If the datagrams make it through the network, they are received by another application. If datagrams are lost, there is no strategy for retransmission. Consequently, there is no guarantee that packets will be in order, and a UDP recipient can deliver data to the application in the wrong order.

Ports are used by UDP to allocate data to the correct application in the destination system, so the port number of the service receiving the data is embedded in the UDP header. UDP also sends a checksum in its header, as a reliable integrity check option, which enables the recipient to detect transmission errors.

UDP does not apply any flow or congestion control to its sending strategy. If applications generate large amounts of data, the network can become flooded. One common scenario is that applications send more data than can be sent on the network, which leads to packet losses for UDP data and possibly for other flows.

From a video surveillance perspective, UDP favors timely delivery of data over reliability, but it does so without the congestion control and burst control that TCP has. UDP tends to be a better choice when trying to minimize delays and jitter. Using UDP may be preferable when transmitting data that requires low latency and which can tolerate some degree of loss, such as multimedia broadcast applications. On the other hand, when bandwidth is limited or if there is a firewall or NAT in the path, then TCP tends to work better.

### 11.4.2 Transmission control protocol

TCP is the most commonly used protocol for data transport and, when used with IP, is often referred to as TCP/IP. TCP divides data into TCP segments for data transmission, adding supplementary flow-control information in the TCP header. TCP provides a connection-oriented, reliable, and

in-order delivery of data streams. In addition, it is responsive to network congestion. These characteristics make TCP suitable for applications such as file transfers or email.

In contrast to UDP, TCP is a connection-oriented protocol. This means that it establishes a connection between two communicating applications before any data exchange takes place. The connection makes sure that data flows only between these two hosts, which, however, prevents the use of TCP for broadcasting or multicasting. TCP also provides transmission reliability: the recipient confirms the incoming data, and, if necessary, the sender retransmits if no confirmation is received.

TCP provides transmission and reception of a reliable and in-order stream of bytes for applications at the upper layers. It does this by splitting the payload into sequentially numbered segments, transmitting those segments according to its protocol rules, and finally, at the receiver, verifying and reassembling the segments to reconstruct the original stream. As data are exchanged, receivers continuously acknowledge successful receipt of segments by issuing sequence numbers so that senders can retransmit lost data as required. If segments are lost, a TCP recipient does not pass out-of-order data to the application. Instead it waits for the retransmitted segments to arrive; that is, TCP introduces a delay in its efforts to be reliable and provide in-order delivery.

While exchanging data, TCP continuously applies flow and congestion control to its sending strategy. For example, when a recipient is slower than the sender, TCP's flow control forces the sender to slow down. Similarly, when the network path is congested, TCP's congestion control also forces the sender to slow down. These are important functions for avoiding congestion collapses, excessive packet loss, and inequalities between flows.

From a video surveillance perspective, conventional wisdom indicates that TCP is not suitable for real-time traffic because it favors reliability over the timely delivery of data. Because of this, many real-time protocols tend to use UDP and fall back on TCP only when absolutely necessary (e.g., due to bandwidth, firewall, or NAT issues along the path). On the other hand, for video applications where delays are not critical—such as for video storage—TCP is commonly used.

## 11.5 APPLICATION LAYER PROTOCOLS

---

On the application layer, which is the highest level of both the OSI model and the TCP/IP model, different protocols are required for data exchange between a network video system and the user, for example, when accessing the menus of a network camera and viewing video. The most common protocols are HTTP, FTP, and Real-Time Transport Protocol (RTP), which are all explained in the following.

### 11.5.1 Hypertext Transfer Protocol

HTTP is used primarily to load the text and images from a website to a web browser. HTTP is a stateless protocol, meaning that a connection is not maintained between systems once data have successfully transmitted. A new connection must be established for additional data transmissions.

Network video systems provide an HTTP server service that permits access to the systems through web browsers, for downloading configurations or live images.

### 11.5.2 File Transfer Protocol

FTP is a network protocol for data transmission via TCP/IP. It is used primarily to transmit files from a server to a client (download) or from a client to a server (upload). FTP also can be used to create and select directories and rename or delete directories and files.

There are two modes for establishing FTP connections: active mode and passive mode. In active mode, the FTP server establishes a connection to the client following a request from the client, whereas in passive mode the client establishes the connection to the server. Passive mode is used if the server cannot reach a client. This is the case, for example, if the client is located behind a router that converts the client's address by means of NAT or if a firewall protects the local network against external access.

Network cameras can use FTP to transmit JPEG images to an FTP server for storage. In such a case, the network camera acts as an FTP client and establishes an event-based connection to the FTP server. It then transmits multiple JPEG images to the server and stores them to a specific directory using different file names.

### 11.5.3 Simple Network Management Protocol

Simple Network Management Protocol (SNMP) represents a set of protocols for managing complex network infrastructures. It can be used to remotely monitor and manage networked equipment such as switches, routers, and network cameras. Many network cameras have support for SNMP, which means they can be managed by tools such as OpenNMS, which is an open source. The latest version of SNMP is version 3.

### 11.5.4 Simple Mail Transfer Protocol

SMTP is the *de facto* standard for transferring email over the internet. Although not normally relevant in network video, many network cameras have support for SMTP to allow the sending of email alerts (e.g., when motion is detected in a scene), which can even include attachments with snapshots or video clips.

### 11.5.5 Real-Time Transport Protocol

RTP permits the transfer of real-time data between system endpoints. RTP is a packet-based protocol that is usually transmitted through UDP. RTP services include identification of the transmitted user data and its sources, as well as the allocation of sequential numbers and timestamps to the data packets. Using this information, the recipient can reassemble the individual data packets in the correct sequence.

Video compressed using H.264 is often transmitted over RTP. In conjunction with RTP, the complementary Real-Time Streaming Protocol can be used for extended control over the transmission of real-time media.

RTP can be used for both unicasting and multicasting applications. See Section 11.6 for more on this.

### 11.5.6 Session Initiation Protocol

Session Initiation Protocol (SIP) is a communication protocol for signaling and controlling multimedia communications sessions. The most important applications of SIP are in internet telephony for voice and video calls, as well as instant messaging over IP networks.

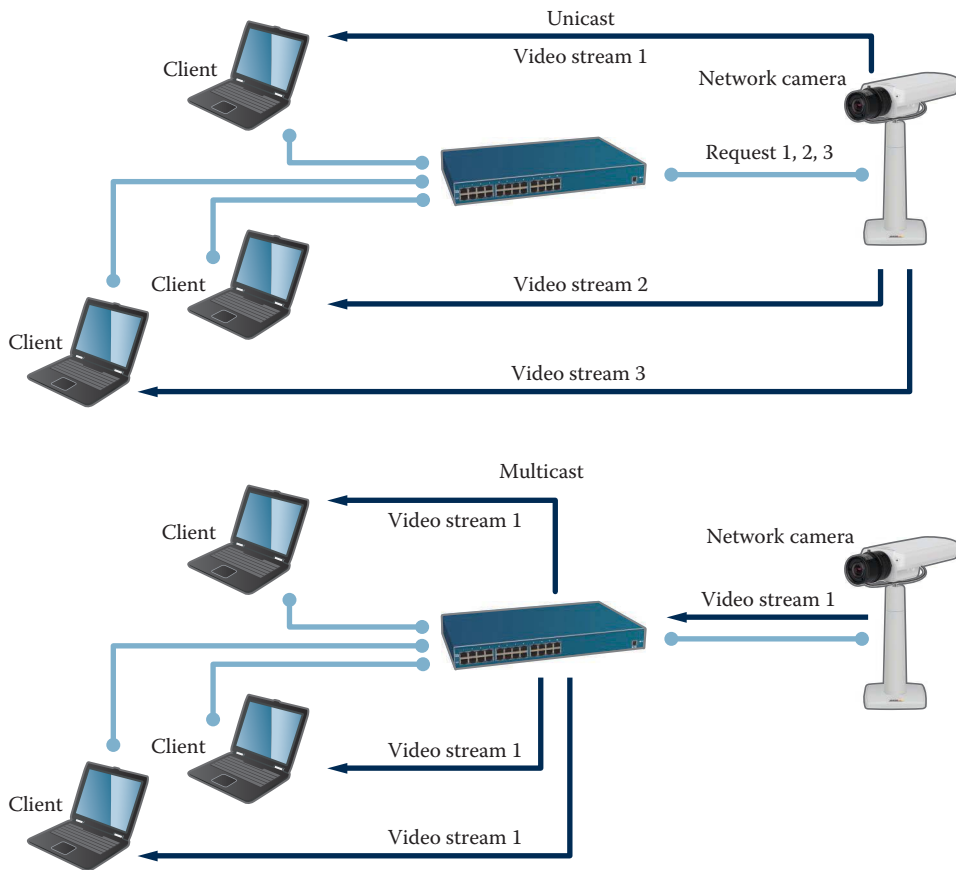
SIP is an application layer protocol designed to be independent of the underlying transport layer. SIP is text based and incorporates many elements of the HTTP and SMTP protocols.

SIP also works alongside several other application layer protocols that identify and carry the session media. Media identification and negotiation is achieved with the Session Description Protocol. For the transmission of media streams (voice, video), SIP typically employs the RTP protocol or the Secure Real-Time Transport Protocol. For the secure transmission of SIP messages, the protocol can be encrypted with Transport Layer Security (TLS) and is then referred to as SIPS. For more information about SIP, see Section 7.8.3.

## 11.6 UNICAST, BROADCAST, AND MULTICAST

There are three different methods for transmitting data on a computer network, each catering to different needs:

1. *Unicast* (Figure 11.12) is the most common form of communication, in which the sender and the recipient communicate on a point-to-point basis. Data packets are sent only to one recipient, and no other clients will receive that information.
2. *Multicast* (Figure 11.12) is the communication between a single sender and multiple receivers on a network. Multicast technologies are used when many clients request the same



**Figure 11.12** Unicast (top) and multicast (bottom) video transmission.

information—for example, live video—simultaneously. Multicasting reduces network traffic by delivering a single stream of information to many recipients. The biggest difference when compared to unicasting is that the video stream only needs to be sent once, whereas for unicast, a copy for each recipient is required. Multicasting is typically used when large numbers of users wish to view live surveillance video. RTP is the most common protocol used for multicasting video streams.

3. **Broadcast** means that the sender sends the same information to all other servers on the network. When a broadcast message is sent, all hosts on the network receive the message and will process it to some extent. Too many broadcast messages will slow down a network and the hosts connected to it. Routers block broadcast messages and are used to create broadcast domains that limit broadcasts to the network segment on which the broadcast originated. In broadcast addressing, a distinction is made between a limited broadcast (address 255.255.255.255), which is not forwarded by routers, and a direct broadcast (e.g., 146.15.255.255), which will be forwarded as necessary by routers so that all hosts are reached. Broadcasts are not practical for network video transmission, and network video products only use broadcasts for specific protocols that require it, such as DHCP.

## 11.7 QUALITY OF SERVICE

At present, fundamentally different applications—for example, telephone, surveillance video, and email—all use the same IP network. In such a network, it is necessary to control how network resources are shared to fulfill the requirements of each service. One solution is to let network

routers and switches operate differently for different kinds of services (voice, data, and video) as traffic passes through the network. Using Quality of Service (QoS), different network applications can coexist on the same network without consuming each other's bandwidth.

### 11.7.1 Definition

The term *Quality of Service* refers to a number of technologies that guarantee a certain quality to different services on a network. Examples of quality include a maintained level of bandwidth, low latency, or no packet losses. The main benefits of a QoS-aware network can be summarized as:

- The ability to prioritize traffic so that critical flows can be served before flows with less priority
- Greater reliability in a network by controlling the amount of bandwidth an application can use and, as a consequence, the ability to control bandwidth competition between applications

QoS relates to the transmission delay (latency) between systems, jitter (variation from the average latency time), packet loss rate (loss probability for individual packets), and data throughput (bandwidth).

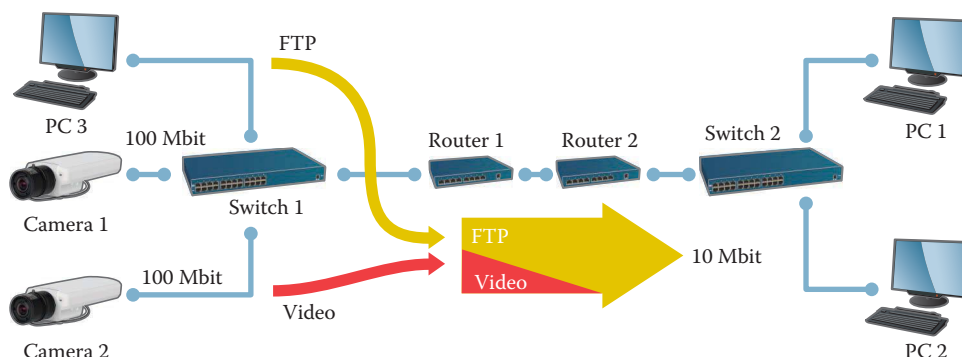
Datagram headers in IPv4 and IPv6 contain a Differentiated Services Code Point (DSCP) flag for identifying the type of data in the relevant IP datagram. Using this flag, the data packets are divided into traffic classes and prioritized for forwarding. The DSCP flag has a length of 6 bits, which means that 64 different classes can be defined.

### 11.7.2 QoS in network video

To use QoS in a network with network video products, the following requirements must be met:

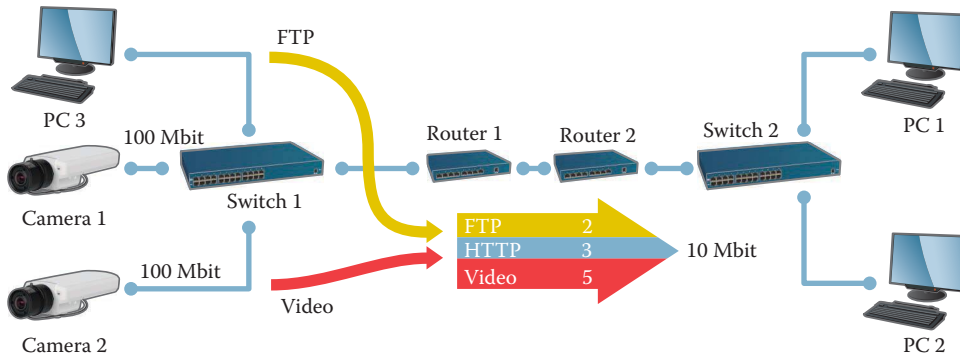
- All network switches and routers must include support for QoS. This is important to achieve end-to-end QoS functionality.
- The network video products must be QoS enabled.

See Figure 11.13 for an ordinary (non-QoS-aware) network and Figure 11.14 for a QoS-aware network.



**Figure 11.13** An ordinary (non-QoS-aware) network. PC1 is accessing video streams at 2.5 Mbit/s from camera 1 and camera 2. Suddenly, PC2 starts a file transfer from PC3. In this scenario, the file transfer tries to use the full 10 Mbit/s between routers 1 and 2, while the video streams try to maintain their total of 5 Mbit/s. The amount of bandwidth for video can no longer be guaranteed, and the frame rate is probably reduced. At worst, the FTP traffic will consume all the available bandwidth.





**Figure 11.14** A QoS-aware network. Router 1 is configured to devote up to 5 Mbit/s of the available 10 Mbit/s to video. FTP may use 2 Mbit/s, and HTTP and all other traffic can use a maximum of 3 Mbit/s. This way, video always has the necessary bandwidth available. Note that these maximums only apply when there is congestion on the network. If there is unused bandwidth available, this can be used by any type of traffic. To guarantee fast responses, PTZ traffic, which requires low latency, is often regarded as critical. This is a typical case where QoS can be used to provide the necessary guarantees.

## 11.8 NETWORK SECURITY

Considering the wide use of IP networks for data, video, and voice by governments, banks, and enterprises, ensuring the secure transfer of information is vital. Although early IP networks had some security flaws, huge investments in research and development have now led to extremely secure networks. There are several ways to provide security both within a network and between different networks and clients. Everything from the data sent over a network to the actual use and accessibility of the network can be controlled and secured.

Secure communication can be divided into two different types:

1. *Authentication and authorization:* This initial step involves the user or device identifying itself to the network and the remote end. This is done by providing some kind of identity—for example, a username and password. This authentication is then authorized and accepted; that is, it is verified whether or not the device is allowed to operate as requested. Once authorization is complete, the device is fully connected and operational in the system.
2. *Privacy:* Privacy is accomplished by encrypting the communication to prevent others from using or reading the data. The use of encryption can slow communications down, depending on the kind of implementation and encryption used. Privacy can be achieved in several ways. Two of the more commonly used methods are virtual private networks (VPNs) and SSL/TLS (also known as HTTPS).

The main technologies for securing data transmissions are explained as follows.

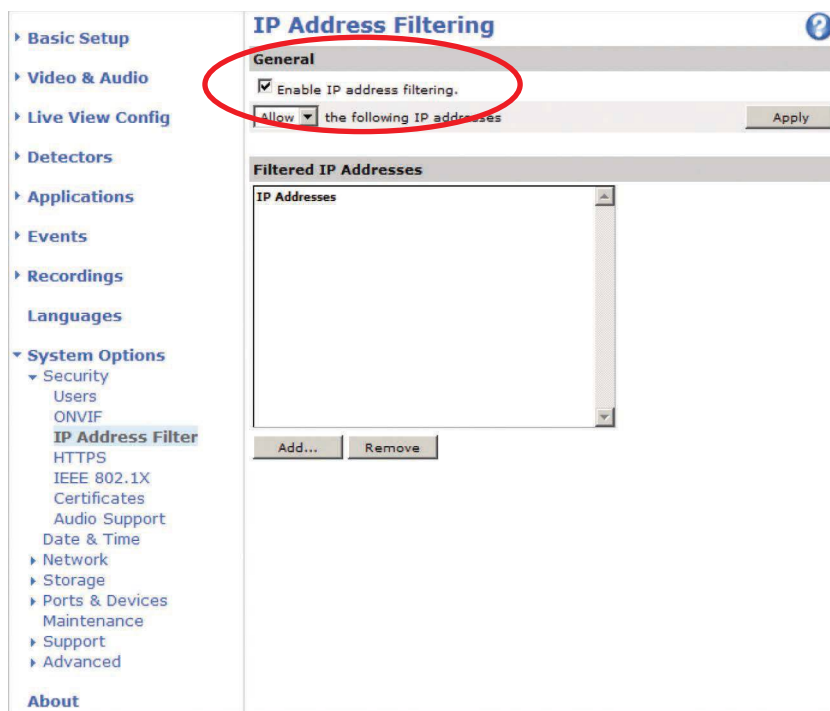
### 11.8.1 Username and password authentication

The most basic method of protecting data on an IP network is to use username and password authentication. Data are protected from access until a user submits a valid username and password. Passwords can be sent as encrypted or unencrypted data, the former providing the better security.

Username and password protection may be appropriate in an installation where high levels of security are not required or where the video network is segmented from the main network and unauthorized users would not have physical access.

### 11.8.2 IP filtering

Many network cameras include IP filtering, which prevents all but one or a few IP addresses from accessing the product (Figure 11.15). It provides a function similar to a built-in firewall.



**Figure 11.15** IP filtering in a network camera restricts access to specified IP addresses only.

Installations that require a higher level of security can use IP address filtering, where a typical setup will configure IP cameras to only allow the IP address of the server running the video management software.

### 11.8.3 802.1X

The IEEE 802.1X standard is among the most popular authentication methods in use today. It provides authentication to devices attached to a LAN port, establishing a point-to-point connection or preventing access from that port if authentication fails. 802.1X is often referred to as Port-Based Network Access Control and prevents port hijacking—that is, when an unauthorized computer obtains access to a network by getting to a network jack inside or outside a building.

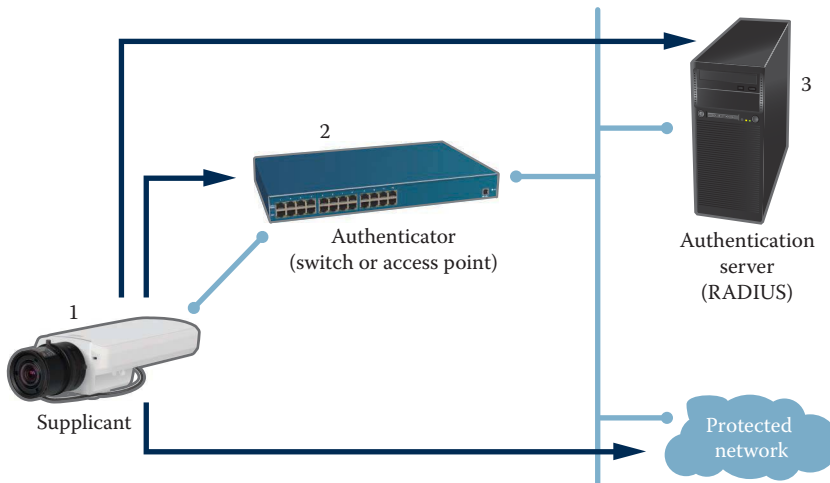
Authentication requires three components:

- The supplicant
- The authenticator and
- An authenticating server

The supplicant is a device such as a network camera that requests access to the network. The authenticator can be a switch or a wireless access point. Logical ports on the authenticator allow user data from the supplicant to pass once the supplicant is authenticated. The authenticating server is a (dedicated) server on the LAN to which supplicants must identify themselves in the authentication process. Some switches today can include the functionality of the authenticating server.

The authenticating server is called a Remote Authentication Dial-In User Service. If a device wishes to access a network, it asks the authenticator, which forwards the query to an authentication server. If authentication is successful, the server instructs the authenticator to authorize access to the network for the querying supplicant (Figure 11.16).

802.1X is often built into network cameras as they are often located in public spaces (such as receptions, hallways, meeting rooms, or even outside a building) where they need the security layer that



**Figure 11.16** IEEE 802.1X enables port-based security and involves a supplicant (1), an authenticator (2), and an authentication server (3).

802.1X provides. Without 802.1X, a network socket that is openly accessible poses a substantial security risk. In today's enterprise networks, 802.1X is becoming a basic requirement for anything connected to the network.

#### 11.8.4 Virtual Private Network

A VPN uses an encryption protocol to provide a secure tunnel, from one network to another, through which data can be securely transferred. This allows for secure communications across a public network, such as the internet. Only devices with the correct “key” will be able to work within the VPN.

A VPN typically encrypts the packets on the IP or TCP/UDP layers and above. The IPSec Protocol is the most commonly used VPN encryption protocol, and it can use various encryption algorithms. Today, either the Triple Data Encryption Standard (3DES) or the Advanced Encryption Standard (AES) is used. AES offers greater security and requires considerably less computing power than 3DES to encrypt and decrypt data. AES uses either 128 or 256-bit key lengths (Figure 11.17).

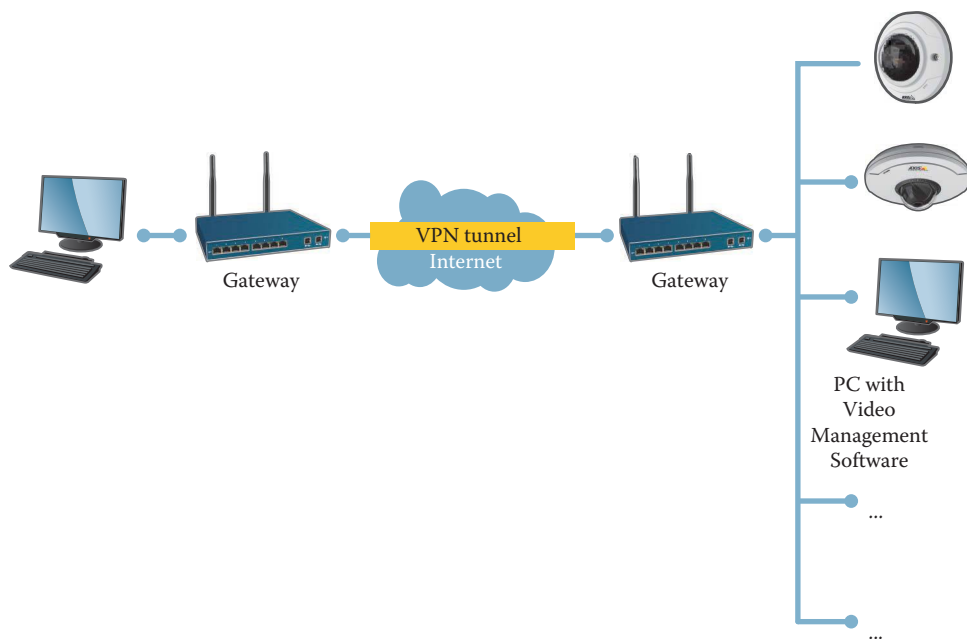
#### 11.8.5 Hypertext Transfer Protocol Secure

Another way to provide privacy is by applying encryption on a higher level—that is, the data but not the transport protocol are encrypted. HTTPS is the most common data encryption protocol. It is commonly used in online banking to provide the requisite security for banking transactions performed over the internet. HTTPS is identical to HTTP, but with a key difference: the data transferred are encrypted using SSL or TLS (Figure 11.18).

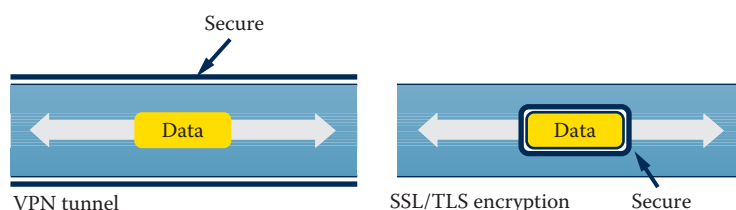
Netscape developed and published SSL in 1994. The security offered by SSL/TLS is based on three main elements:

- Authentication of the communication partner
- Symmetrical data encryption and
- Protection against the manipulation of transferred data

When making an SSL/TLS connection, negotiation first takes place through the use of a handshake protocol that determines which cryptographic methods will be used, as well as the secure identification or authentication of the communication partner. This last element is achieved by the web server



**Figure 11.17** Virtual private network security on a video network.



**Figure 11.18** The difference between VPN and SSL is that in a VPN, the tunnel itself is encrypted, whereas in SSL, the data are encrypted. Both technologies can be used in parallel, but this is not recommended because each will add an overhead and affect the performance of the system.

identifying itself to the browser using a *certificate*. A premaster secret is then exchanged between the communication partners over an asymmetrical encryption or Diffie–Hellman key exchange.

A certificate can be compared to an ID card that a person uses to prove his or her identity. This is a binary document usually issued by a Certificate Authority, such as VeriSign. Certificates normally used only within closed user groups, such as for a company’s private web server, can be issued by anyone.

Many network video products have built-in support for HTTPS, which makes it possible to view video securely using a web browser. However, the use of HTTPS can slow the communication link down and thus lower the video frame rate.

## 11.9 BEST PRACTICES FOR NETWORK SECURITY

All network devices are subject to threats. This includes network cameras, which are always part of a larger system where the network is the backbone. All parts are vulnerable, either as a system or as individual devices, and the system needs protection. It is not, however, possible to create a system that is 100% secure, at least not a usable system. The system can only be made more secure by reducing exposure areas and mitigating risks, but there will always be some level of risk that needs to be accepted.

A cybersecurity vulnerability is a weakness in firmware, hardware, system interfaces, system services, software, or integration that exposes a flaw that can be exploited for a malicious attack. This does not mean that someone would easily be able to exploit such a weakness. In many cases, an attack requires various conditions to first be fulfilled, such as access to the network and its resources.

### 11.9.1 Hardening guides

Many manufacturers have released recommendations on how to ensure that appropriate network security is deployed in a network video system. Such recommendations are often referred to as hardening guides, and they refer to different levels of protection, examples of which are given here:

- *Default protection (Level 0)*: Cameras are shipped with preconfigured default settings and a default password. Default settings should not be used for daily operations. Settings must be adjusted depending on the environment and risk analysis.
- *Standard protection (Level 1)*: This is the minimum recommended level and is sufficient for home and small office use, typically when the system administrator is also the operator/user. Other actions include checking for the latest firmware, setting the time and date, and disabling audio if not in use.
- *Enterprise protection (Level 2)*: Corporate and enterprise system have dedicated system administrators and the network has multiple clients (users) and servers. Video systems must adhere to the corporate IT infrastructure and policies. Security capabilities and features must be enabled and configured, whereas unused camera services should be disabled. Other actions include enabling IP address filtering and encryption.
- *Managed enterprise (Level 3)*: Managed enterprise network system will typically have additional management tools and services that cameras must align with.
  - 802.1X Network Access Protection
  - VLAN segmentation
  - SNMP Monitoring

### 11.9.2 Best practices

- *Risk analysis*: Try to determine potential threats and the possible damages and costs if the system is attacked. This will indicate the level of protection required, as well as the precautions that must be taken to reduce the risks.
  - Evaluation and assessment—identify assets and evaluate their properties and characteristics
  - Risk assessment—discover threats and vulnerabilities that pose a risk to assets
  - Risk mitigation—address a risk by transferring, eliminating, or accepting it
- *Educate yourself*: Gain knowledge of system protection and possible threats. Work closely with resellers, system integrators, consultants, and product vendors. The internet is a fantastic resource.
- *Secure the network*: If network protection is breached, this increases the risk of snooping for sensitive information and attacks on individual servers and network devices.
- *Passwords*: Use strong, unique passwords and change them on a regular basis.
- *Factory default settings*: Do not rely solely on a network device's default settings.
- *Encrypted connections*: Use whenever possible, even on local networks.
- *Reduce exposure*: Video clients should not access cameras directly, unless required. Clients should normally only access video through a video management system or a media proxy.
- *Check logs*: Check access logs on a regular basis to detect attempts at unauthorized access.

- *Monitor devices:* Monitor devices on a regular basis. Enable system notifications where applicable and supported.
- *Firmware:* Use the latest available version, which may include security patches.

While cybersecurity is of the utmost importance, remember that the greatest threats to a camera are physical sabotage, vandalism, and tampering. Selecting a vandal-resistant model/casing and mounting it correctly is important, as is protecting cables.



# CHAPTER 12

## Servers and storage

Servers and storage are important components of a network video system as they are used for monitoring, recording, managing, and archiving video. A major benefit of IP-based video surveillance is the ability to deploy standard server and storage components from the IT world. Components can range from a low-price desktop PC to a blade server system attached to a storage area network (SAN) with petabytes of storage that can cost millions of dollars. As they are easy to use and install, purpose-built network video recorders (NVRs) are popular in small and some midsize systems, especially if the NVRs have built in PoE ports to power the cameras. Some of those NVRs are proprietary in nature and are not based on standard servers and components.

Another benefit of using standard components is that they can be serviced and maintained along with the other IT equipment within an organization. This not only reduces costs but also increases the availability and reliability of the equipment.

Because all the equipment consists of standard IT components, this chapter provides just a brief overview of server and storage technologies, along with some best practices relating to network video.

### 12.1 SERVERS

---

Depending on the performance required, there are various aspects to consider when selecting the appropriate server platform. The rule of thumb in the IT industry is that processor chip performance at a certain price level doubles every 18 months (a.k.a. Moore's law). For a network video system, the benefit of fast development is that servers are easily upgraded. For example, if greater performance and capabilities are required, then the video management software can be installed on a new server.

#### 12.1.1 Hardware platforms

Most hardware platforms are based on Intel® or AMD (Advanced Micro Devices, Inc.) processors. There are, however, some differences between a desktop platform used for word processing and email and the enterprise-class servers normally used in network video applications. Enterprise servers are designed for handling huge amounts of data and for continuous reading and writing. Thus, they differ to a desktop PC in the following areas:

- *Processors:* Compared to a PC with a single processor, an enterprise server is usually equipped with a multicore processor or multiple processors.
- *Internal memory:* The internal random access memory (RAM) is the memory used for processing and buffering data. Because a server in a video system must handle large amounts of data, large RAM buffers are required. Even the smallest enterprise server will have 8–16 GB (giga-bytes) RAM, and a top-of-the-line server may have several hundred GB.

- **Storage:** Compared with a standard PC designed for occasional reading and writing, enterprise systems require reliable storage solutions with high throughput and continuous reading and writing so that they can serve multiple clients. Specific hard drives rated for video surveillance data (with continuous writing) are also available. Note also that neither the operating system (OS) nor the video management system should be run on the same disks as those used for video storage.

There are also other server parameters to consider. Servers are available in various designs, such as stand-alone towers, rack-mounted servers, or blade servers. Blade servers are popular because of their high density and because they are suitable for larger video management systems. Virtualization is another popular technology, where one physical server runs several virtualized servers on the same hardware, although this type of setup is not generally used for video storage or display.

### 12.1.2 Operating systems

Making the server hardware operational requires an OS. In theory, any OS can be used as a platform for recording and managing video, as long as the video management software supports the OS. The choice of an OS depends on a number of technical requirements. To streamline administration and management of their business applications, most organizations standardize on a single OS platform. This choice may also dictate the choice of video management platform. The most commonly used platforms are Windows® and Linux®:

- **Windows:** Windows is the most common platform for video management applications. Using Active Directory® service, it is possible to provide centralized authentication and authorization services for Windows-based computers. The Windows file system is New Technology File System (NTFS), which supports a file system up to 256 TB (terabytes) and file sizes of 16 TB.
- **Linux:** Linux is a popular UNIX™-like OS that comes in a variety of distributions. It is not commonly used in network video applications, but is very popular as an embedded platform for network video devices such as network cameras and video encoders. The most common file system used is ext3 (third extended), which supports file systems up to 35.2 TB and file sizes up to 2.2 TB.

### 12.1.3 Video file systems

As discussed earlier, different file systems have different capabilities. Regardless of the OS or file system, it is important to have a well-organized structure when storing recorded video on a hard disk. Video can be stored as standard files (JPEG, MPEG, H.264, AVI, ASF, etc.) or as raw indexed data. The structure can be a simple directory tree (with subdirectories containing files for each camera, week, day, and hour), or it can be a complex database (with or without multiple references) together with metadata information.

There is no standardized method of organizing large amounts of recordings. Video management software vendors use different technologies; some use standard databases such as Structured Query Language (SQL) or Oracle, or file systems, and others use proprietary formats and structures. Even two versions of the same application can differ in terms of how video is stored. The structure can be optimized for high recording performance, fast searches, tampering detection, stability, and recovery, among others. Unless an application has an open interface for exporting, accessing, or searching stored video, it can be extremely difficult for other applications that use a different system to read and interpret the recordings.

Before video is even stored to disk, it is important to consider the properties of the video stream being saved. It can be well worth the effort to carefully select compression, frame rate, and other video quality-related parameters since they have a great impact on the amount of storage. For more on video compression, see Chapter 6.

## 12.2 HARD DISKS

The hard disks used with PC servers and desktops today range from 1 up to 16 TB and increase with greater capacity every year. In addition to size, there are several other factors that differentiate hard disks. One is the spin speed, measured in rotations per minute (RPM), which for consumer-grade products is normally 5400 or 7200 RPM. Another is the nonsequential (random) read-write performance of the disk system, which indicates the speed of reading and writing to nonsequential blocks of data.

Of particular interest to network video systems are surveillance-grade hard disks. In contrast to a standard desktop or server hard disk, a disk of this type is designed for almost constant write operations. As there usually is far more saved video than is ever viewed, the disk's read-write head can move a little slower and more smoothly. This radically decreases the level of mechanical wear. These disks are usually also much better at power management and are better equipped to withstand the extremes of temperature associated with constant write operations.

Hard disks are supplied in various form factors. For servers, a 3.5-in. standard casing is the norm. Another major differentiator is the interface, with some of the most common interfaces described in the next section.

### 12.2.1 Small Computer System Interface

The Small Computer System Interface (SCSI) is a standard for physically connecting computers and peripheral devices such as hard disks, tape drives, and optical drives. The original interface built on parallel communication, whereas most modern SCSI interfaces transfer data via serial communication, which gives faster data rates, longer reach, and better cabling options.

SCSI hard drives are good choices for demanding video surveillance systems. They are designed and optimized for server applications require high performance and durability and for continuous reading and writing 24/7. Table 12.1 provides an overview of some of the many SCSI variants.

### 12.2.2 Advanced Technology Attachment and Serial Advanced Technology Attachment interfaces

Advanced Technology Attachment (ATA), also known as IDE, ATAPI, and PATA, is another standard for physically connecting computers and peripheral devices such as hard disks, tape drives, and optical drives. Up to two devices can connect to each controller (master and slave). The transfer rate can be as high as 80 MB/s, but 66 MB/s is more realistic. These drives were designed for workstations and laptops for sporadic reading and writing.

**Table 12.1** Some of the many SCSI variants developed over the years

Interface	Type	MB/s	Range	Devices
SCSI-1	Parallel	5	20 ft (6 m)	8
Fast SCSI	Parallel	10	10 ft (3 m)	8
Fast-Wide SCSI	Parallel	20	20 ft (6 m)	16
Ultra-320 SCSI	Parallel	320	39 ft (12 m)	16
Serial Storage Architecture (400 MHz)	Serial	80	82 ft (25 m)	96
Fiber Channel 8 Gbit	Serial	788	6.2 miles (10 km)	2 <sup>24a</sup>
SAS 1.1 (Serial-Attached SCSI)	Serial	300	20 ft (6 m)	16,256
USB-Attached SCSI	Serial	~1200	10 ft (3 m)	127
iSCSI	Serial	Implementation and network dependent		2 <sup>128b</sup>

<sup>a</sup> Theoretical limit.

<sup>b</sup> For IPv6.

ATA is superseded by Serial Advanced Technology Attachment (SATA), which is now the primary architecture in the PC and laptop market. SATA-600 drives have a performance similar to the highest-performing SCSI. The MTBF for SATA drives is similar to ATA, that is, 150,000–800,000 hours, which does not prevent their use in mission-critical installations, as long as they are in a redundant array of independent disks (RAID) configuration (see Section 12.4.1).

Serial-Attached SCSI (SAS) is a serial communication protocol that allows for much higher speed data transfers, with up to 128 direct point-to-point connections. As SAS devices are compatible with SATA, this means that support for SCSI devices can be included on, for example, a motherboard at nominal extra cost. Furthermore, the fact that SAS devices use serial communication means they are not exposed to the interference (cross talk) that occurs over parallel communication. Maximum cable lengths for serial communication are generally longer than for parallel devices.

### 12.2.3 Hard disk failure rates

Although failure rates for today's mechanical hard drives are much lower than for previous-generation products, according to research (reviews of drive vendors' MTBFs and third-party hard disk drive reliability studies), the most common failures in the current generation of hard drives appear to be related to the head-disk interface. These problems can have many different sources, including handling damage, temporary interface disruption, media damage, and thermomechanical stability of the read and write structures. Regarding the timing associated with an actual hard drive failure, research suggests a failure is likely to occur in either the first 60 days of service, or in the third, fourth, or fifth year of operation, given the high duty cycle and operational characteristics expected in a video surveillance storage application.

Because data loss can be very costly, using RAID configurations for redundancy in video management system is recommended. A new server in a video surveillance system should be subjected to a "burn-in" test to ensure the drives do not fail within the first few days.

### 12.2.4 Solid-state drives

Unlike the mechanical disks discussed earlier, a solid-state drive (SSD) is a storage device without moving parts or disks. Instead, it uses nonvolatile flash memory for its data storage, thus endowing the SSD with various desirable properties, including silent operation, fast access, low latency, and good resistance to shocks and vibrations.

Although fast and reliable, SSDs are considerably more expensive per GB of storage than traditional mechanical disks. Device capacity can also be a limiting factor when considering storage for a network video system, as there are few SSD drives that offer the same storage capacities as mechanical drives.

## 12.3 STORAGE ARCHITECTURE

---

The demand for more and more storage for all kinds of applications is driving the industry to develop ever-better performance and greater capacity for storage systems. High-quality surveillance video puts very high demands on a storage system, and storage is often a considerable part of the cost of a network video system. Virtually any size of storage system can be accommodated, which means that any frame rate, number of cameras, and retention time can be handled.

There are a few basic types of architecture for storage systems, all with different complexity, performance, cost, redundancy, and scalability. The most common are described in the following sections.

### 12.3.1 Edge storage

In some applications, storing the video onboard—that is, in the network camera itself—can be a good choice. More and more cameras provide a slot for a standard memory card such as an SD card,

which can allow many days' of video to be stored, all depending on the frame rate, the resolution, and the capacity of the SD card. In small systems, using SD cards as the only type of video recording is today feasible and provides for a very cost-efficient solution.

As mentioned previously for mechanical hard drives, the saving of video to a physical storage medium is a very write-intensive process. For this reason, it is important to always use surveillance-grade SD cards for edge storage applications. These SD cards are specially designed to handle the large number of write operations typical when saving video and will last much longer than a standard SD card.

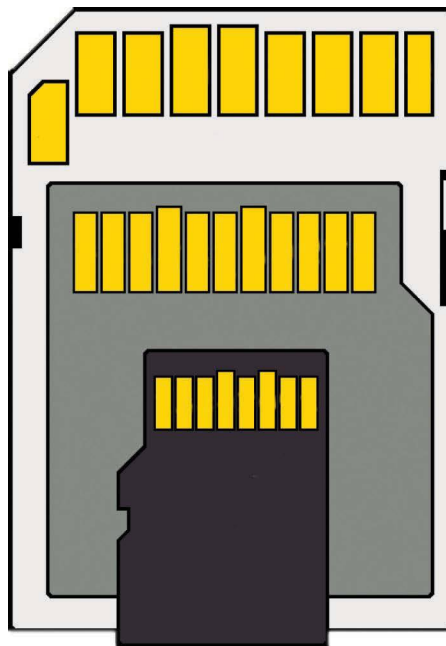
Another scenario for a distributed storage model is a mission-critical one where loss of video is not acceptable when the network goes down or when taken offline for maintenance. This scenario is more hypothetical in nature because today's networks have a very high degree of reliability, provided that they are designed correctly.

The term *SD card* is used to cover three different formats of SD card:

- SDSC (Standard Capacity) is the original format and supports up to 2 GB of storage
- SDHC (High Capacity) supports up to 32 GB
- SDXC (eXtended Capacity) supports up to 2 TB

Also of interest is the speed class of the SD card, which specifies the minimum write speed. In a network video system, this factor is very important, as selecting the correct speed class guarantees that video can be written to the SD card quickly enough. Write speeds range from 2 MB/s for Class 2 to 10 MB/s for Class 10 sufficient for 1080p video, all the way up to 30 MB/s for UHS Speed Class 3, which is sufficient for saving video at 4K resolution.

In parallel with the advances in storage capacity and write speed, the form factor of SD cards has also changed. Two steps down in size from the original SD card have produced the miniSD and the microSD, as shown in Figure 12.1.



**Figure 12.1** A comparison of the sizes of the original SD card (largest), the miniSD card, and the microSD card (smallest).

### 12.3.2 Single-server storage

Single-server storage is a possible solution for video storage in a small installation of up to 16 cameras (Figure 12.2). The hard disks are located in the same PC server that runs the video management software (application server). The PC and the number of disks it can hold will determine the available space. Most modern PCs provide support for multiple hard drives, which with terabytes of space can be more than enough for the most demanding application, even with long retention times.

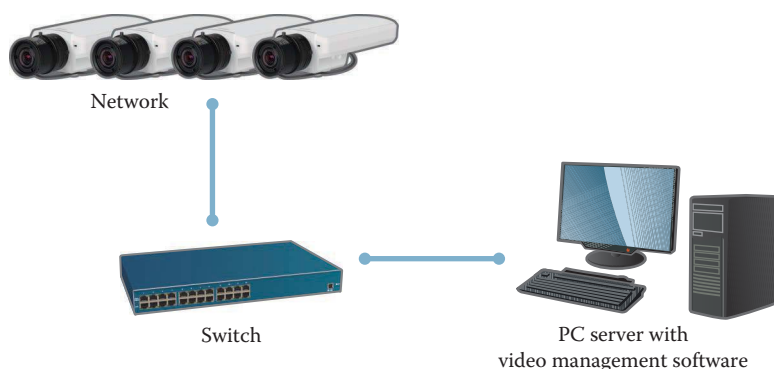
In applications where the amount of stored data and management requirements exceed the limitations of direct attached storage, a separate storage system must be implemented. These systems are network-attached storage (NAS) and SAN.

### 12.3.3 Network-attached storage

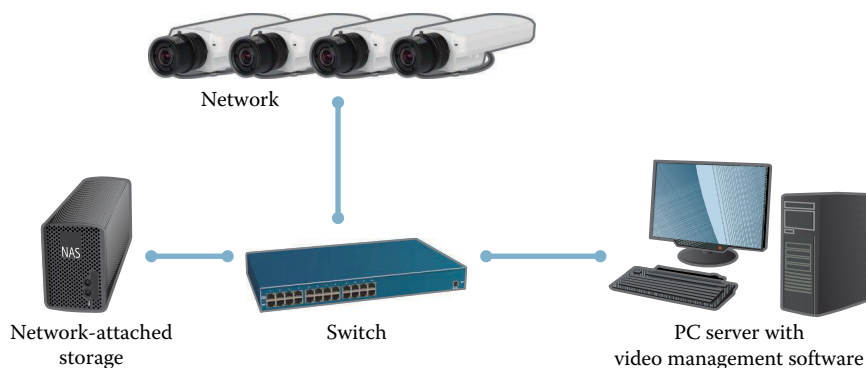
NAS provides a single storage device directly attached to a local area network, and it offers shared storage to all clients on the network. A NAS device (Figure 12.3) is simple to install and easy to administer. It provides a low-cost solution for storage requirements and a good solution for smaller systems.

### 12.3.4 Storage area network

A SAN is a high-speed, special-purpose network for storage devices. It is connected to one or more servers via a fiber channel (Figure 12.4). Users can access any of the storage devices on the SAN through the servers, and storage is scalable to hundreds of terabytes or even petabytes (1000 TB).

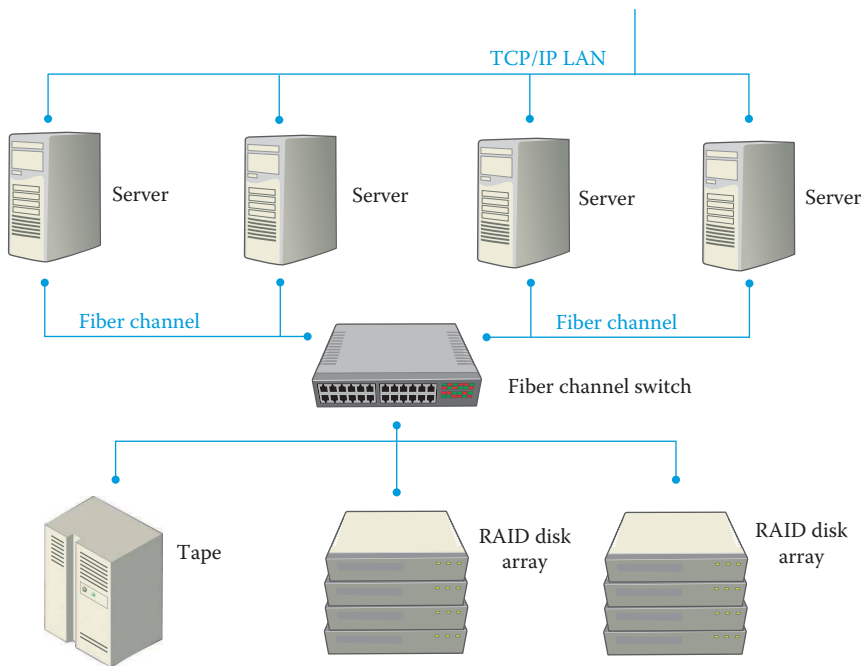


**Figure 12.2** An example of a single-server storage setup.



**Figure 12.3** An example of a network-attached storage setup.





**Figure 12.4** Typical storage area network architecture, where a fiber channel network ties all storage devices together and lets the servers share the storage capacity.

Centralized data storage reduces the administration required and provides a flexible, high-performance storage pool for use in multiserver environments.

The difference between NAS and SAN is that NAS is a storage device in which the entire file is stored on a single hard disk, whereas a SAN consists of a number of devices where the file can be stored, block by block, over multiple hard disks. This type of configuration allows for a very large and scalable hard disk solution, in which vast amounts of data can be stored with a high level of redundancy.

### 12.3.5 Internet Small Computer System Interface

Internet SCSI (iSCSI) is a network protocol for sending SCSI commands over a TCP/IP network instead of an SCSI cable. This makes it possible to connect a storage device and address it as a local drive on the server. Unlike NAS, iSCSI devices do not support file system protocols. The main difference between iSCSI and NAS is that iSCSI provides block-level disk access, whereas NAS only provides file-level access. The host server manages the file system as it would on a local drive. iSCSI is commonly used in large installations, and usually in RAID configurations, as it is a cost-efficient way of quickly adding more storage to a server.

## 12.4 REDUNDANCY

The server and storage components are essential in any IT system, and an IP-based video surveillance system is no different. Many different technologies are available to make the system more reliable by increasing the redundancy. The following sections discuss the most common techniques.

### 12.4.1 Redundant array of independent disks systems

RAID, previously called redundant array of inexpensive disks, is a method of arranging standard, off-the-shelf hard drives in such a way that the OS sees them as a single large logical hard disk.

Benefits include increased throughput and better reliability. Using hardware RAID controllers instead of software controllers is recommended to reduce performance issues.

There are different levels of RAID that offer different levels of redundancy—from practically no redundancy at all to a fully “hot swappable” mirrored solution, in which there is no disruption to operations and no loss of data in the event of a hard disk failure.

The most common RAID levels include the following:

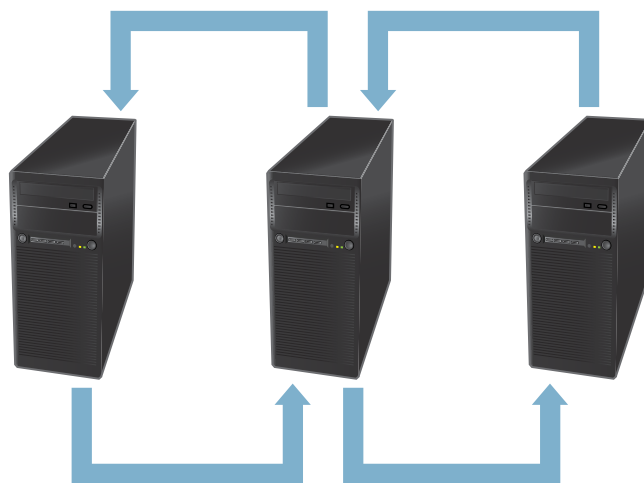
- **RAID-0:** Also called striping. Information is spread over two or more disks to increase performance. There is no redundancy because the array is ruined if one disk fails.
- **RAID-1:** Also known as disk mirroring. The information on one disk is duplicated on one or more other disks. This increases reliability but may reduce performance as data need to be written to multiple disks.
- **RAID-5:** Also known as striping with parity. Data and parity are spread over three or more disks and require at least three disks in the array. Read performance is the same as for a single disk. Write performance can be lower because data must be written to two disks. RAID 5 can tolerate a single disk failure and still recover all data. Additionally, the disks can be made hot swappable. RAID-5 has become popular because it provides redundancy and maximizes disk space for data instead of backup.
- **RAID-6:** Similar to RAID 5 but with dual parity bits. This requires at least four disks, and the configuration can tolerate two disk failures.

### 12.4.2 Data replication

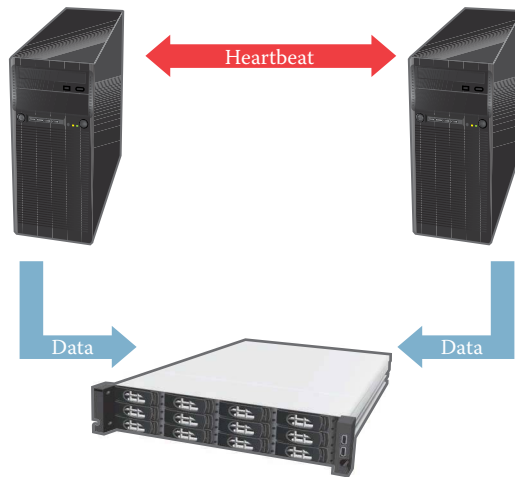
Data replication is a common feature of many network OSs; file servers are configured to replicate data from other servers (Figure 12.5).

### 12.4.3 Tape backup

Tape backup is an alternative or complementary method. There is a variety of software and hardware equipment available on the market. And as prevention against fire or theft, backup policies normally include taking tapes offsite. However, due to the relatively slow write speed of tape drives, these devices are seldom used for backing up video material.



**Figure 12.5** Replicating data ensures a high level of redundancy.



**Figure 12.6** Server clustering is a popular way to ensure redundancy on the server part of the system.

### 12.4.4 Server clustering

Many server clustering methods exist. A common scenario for database servers and mail servers is for two servers working with the same storage solution, commonly a RAID system. In such a case, when one server fails, the other (configured identically) takes over the application (Figure 12.6). These servers usually even share the same IP address, making the so-called failover completely transparent for the user.

### 12.4.5 Multiple servers

A common method to ensure disaster recovery and off-site storage of network video is to simultaneously transmit video to two different servers located in separate locations. In turn, these servers can be equipped with RAID and work in clusters, or they can replicate their data to servers even further away.

## 12.5 BEST PRACTICES

Selecting the correct server and storage platform for a video surveillance system requires many considerations. Some of the questions that need addressing include the following:

- *Has the organization standardized?* In most organizations, the type and brand of server and OS are fixed, so the video management system should also be run on the same type of server. The obvious benefit is that service and maintenance agreements are in place, and the IT department can manage the video management server just like any other server.
- *Is a centralized or decentralized system preferred?* This depends on the size of the system and the available bandwidth between the different locations.
- *What is the required system reliability?* The higher the requirements, the better the grade of disk and server required. Consider investing in surveillance-grade drives.
- *What is the required system redundancy?* A RAID-based recording system might be a smart investment that ensures no video is lost if a hard disk fails.
- *What is the size and scalability?* How many cameras will be managed, and how scalable does the system need to be? Always plan for growth. The server should not be used at maximum performance, and it should not be the case that 90% of the storage space is used right from day one. There will always be a need to add to the system down the road, so plan for such a system today.

- *What is actually being recorded?* With the correct configuration, it does not necessarily need to be the case that doubling the frame rate also doubles the required disk space. Other methods can also lower requirements, for example, using motion-based recording at night and reducing the resolution where possible. See Chapter 6 for more on video compression.
- *How long will video be retained?* Going from 15 to 30 days will, however, double the storage requirement. Remember that more than 99% of all video is never watched. Having different retention times for different cameras is recommended, that is, a few days' retention for some cameras and longer retention times for other more important camera locations.

## CHAPTER 13

# Video management

Video management is one of the major keys to a cost-effective and successful video surveillance system. Selecting the right video management system requires much consideration, including strategies for scalability, flexibility, and functionality. Along with cost, bandwidth, and storage, you need to consider the different management architectures and how they serve different user types. A well-designed system should also be easy to set up and easy to use.

Video management systems come with recording and viewing functionalities as well as functionalities for finding and configuring network video devices. They also include event handling and system security. More advanced functions include video motion detection (VMD) and other video analytics as well as integration with other systems such as physical access control and building management. Video management system and video management software are both abbreviated VMS. This is problematic because people in the industry will sometimes use either of the three terms to mean only the software, whereas others use the same term to mean the whole solution (or parts of it) of integrated hardware, software, and storage. Depending on the context and level of clarity required, you will in this book see several synonyms, such as video management system, VMS, and VMS system; video management software, VMS, VMS software, and VMS application.

Video management software is available for all system sizes:

- Large systems (>100 network video devices, sometimes thousands)
- Midsize systems (10–100 network video devices)
- Small systems (1–10 network video devices)

However, system size is only one aspect to consider when selecting a video management system. On a higher level, video surveillance needs to meet certain levels of comfort, forensic evidence, and preventive ability. The security demands and monitoring activity levels are in many ways more important than the amount of cameras when evaluating and choosing a software, system setup, and platform. Various departments and businesses use the video management system differently depending on their purposes and needs. For example, in a grocery store, one individual might use the video surveillance system for security purposes, whereas another might use it for studying customer traffic patterns.

Monitoring activity can be divided into the following levels:

- *Passive*: Surveillance to remedy feelings of curiosity or insecurity. Monitoring is casual and typically only happens randomly or when following up on alarms. This monitoring level is common in homeowners and small offices.
- *Occasionally active*: Surveillance to remedy feelings of curiosity or insecurity. Monitoring activities are likely to be regular, but they are usually not frequent enough to be fully preventive.

This monitoring level is common in retail businesses and office buildings. This monitoring level is common in retail businesses and office buildings.

- *Active:* Surveillance to prevent incidents. Personnel actively monitors the video and acts on alarms. Monitoring can also have nonsecurity purposes, such as predicting and tracking customer behavior. This monitoring level is common in retail businesses, banks, office buildings, and cities.
- *High security:* Surveillance for critical infrastructure and vital operations. Dedicated security personnel monitors the video 24/7 and acts instantly on incidents. This monitoring level is common in airports, hospitals, metro systems, and prisons.

This chapter provides an overview of the different types of video management systems, the platforms, and the most common functionalities available, in addition to some examples of integrated systems.

## 13.1 VIDEO MANAGEMENT ARCHITECTURES

There are three types of platforms for network video management systems:

1. *Server-based video management:* A central server with software or an appliance manages the video from cameras and encoders. Server-based video management can be divided further, usually into these two categories:
  - a. *PC server with video management software (VMS)*
  - b. *Network video recorder (NVR)*, which is a type of network appliance
2. *Edge-based video management:* The cameras or video encoders manage the video individually. A client application presents the information in one user interface, which makes the system look like a complete system of several devices.
3. *Cloud-based video management:* A software application that runs on servers in the cloud manages the video from cameras and encoders. Cloud-based video management is also known as hosted video.

Each platform has its advantages and disadvantages. In some systems, it makes sense to combine different solutions. In large operations, there might be hundreds of cameras at the headquarters, in which case it makes sense to manage them in a server-based system with PC servers and a VMS. In smaller remote locations with 10–20 cameras, appliances such as NVRs might be easier to install and manage. In very small remote locations with less than 10 cameras, where a few cameras store video on SD cards, edge-based video management might be the most efficient solution. Then, one or two cameras in each location can act as backup cameras that keep video available for management in a cloud-based application.

### 13.1.1 Server-based video management

In a server-based architecture there is a central server, or appliance, that manages the configuration of the system and the video from each camera or video encoder. Server hard drives are usually the primary storage point, but other types of storage devices can also be used. Network-attached storage (NAS) devices can function as main storage or backup storage. Some cameras and encoders have SD card slots for onboard storage. In a server-based system, you would typically use SD cards for backup video recording, so even if the network is down and you lose the connection to the network servers, you can still record video. When the network is up again, the video on the SD cards moves to the server. This is sometimes called video trickling. This use of SD cards or NAS should not be confused with edge-based video management, where the cameras and encoders actually manage video independently of any central servers.

Server-based video management architecture can be divided into two categories: PC-based systems with VMS software, which is typically used in scalable systems with 20 cameras or more, and NVR



systems, which are suitable for 5–20 cameras. In NVRs, the software comes preinstalled with the hardware, which means the systems are easy to deploy, but they do not scale well.

### 13.1.1.1 PC server with VMS software

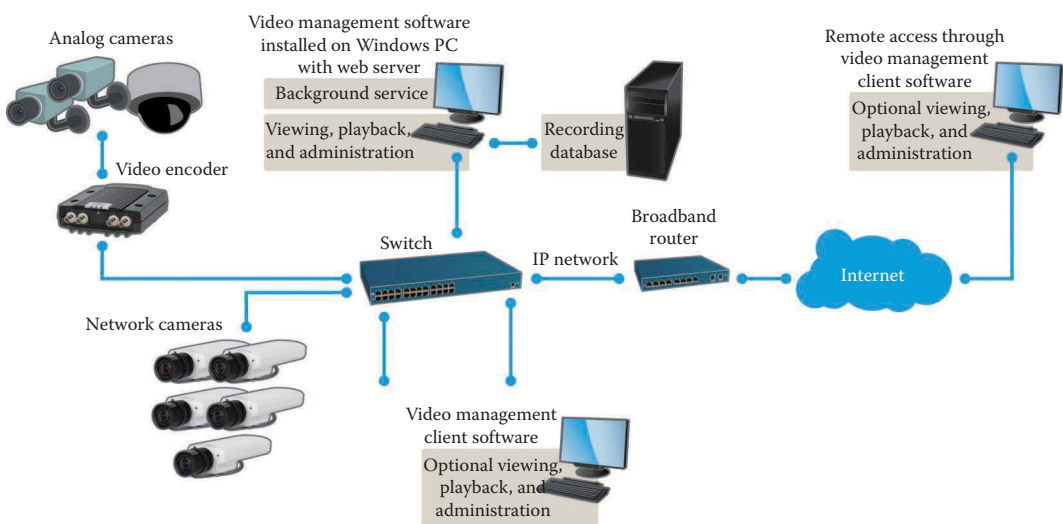
PC server platform solutions are based on commercial off-the-shelf hardware. The hardware components, such as multiple-processor systems and storage, can be selected to obtain the maximum performance for the specific design of the system. A PC server platform solution can use standard components for increased or external storage, for additional remote operator stations, and for running additional software, such as firewalls and virus protection, in parallel with the video application.

Usually, a Windows® platform or a UNIX®/Linux® platform serves as the basis for these systems, although there are also some that run on OS X®. Often a system integrator is responsible for implementing the system and installing the software on the server, but users with the right skills could install the software themselves.

Systems designed on a PC server platform are easily scalable (see Figure 13.1). The system hardware and software licenses can be expanded or upgraded to meet increased performance requirements. The PC server platform is suitable for system scenarios when deploying large numbers of cameras or when the IT department has standard specifications on the type of server hardware and software to allow them on the network. In demanding environments, the video management software can be installed on a ruggedized, industrial PC server.

Often, the viewing client software is installed on a computer that is separate from the recording server where the management software is installed. The server shares all the video management settings with the client, and the client's user interface looks the same and provides the same actions and options. In some cases, users of the client software can switch between different video management servers. This makes it possible to manage video at many remote sites or in a large system.

Video management software is available for all system sizes and monitoring activity levels. The choice of system also depends on camera compatibility and integration with other systems. Some software work best when the camera count stays within a certain range. Normally, a one-camera-one-license principle applies to the video management software, and licenses can be added one by one. System designers have to look carefully at the scalability of the system and factor in future camera counts, user licenses, and number of locations. They also need to consider the activity level



**Figure 13.1** A network video system with server-based video management. The VMS software, which manages the recordings and data, runs on a PC server. The client version of the VMS software, used for viewing live and recorded video, runs on regular computers.

and purpose. Will there always be someone monitoring the live video or is the monitoring activity irregular and alarm based? Is the ambition to prevent incidents or only investigate them when they already have occurred?

Software intended for large or high-security systems often provide full integration with other systems such as physical access control systems (PACSS) or building management systems (BMSs). More often than software for midsize systems, they are open systems that support more than one camera brand. In addition, they offer simultaneous access to multiple sites with advanced functionality. This means that companies that have many sites can streamline their video surveillance without sacrificing efficiency or security. There is great diversity in this group of software. Some are basic and others are feature rich, some are segment specific, and others are complex enterprise solutions.

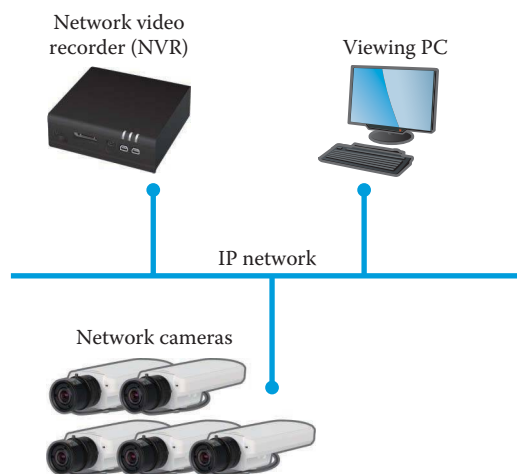
Most software intended for small and midsize systems has functionality that seems more basic compared to what users of large systems are accustomed to. Some of these support only one camera brand and usually have less opportunities for integration with other systems. As a result, they are mainly used for video surveillance. In return, they are generally quite feature rich to support active, targeted, and frequent usage. Typically, when vendors release new products, they update the software to support the features of the new product. In other words, this type of software is perfect for homogenous systems because its features align so well with the product's features.

The main type of storage used in PC server systems with management software is dedicated storage servers, although many support failover recording to edge storage devices such as SD cards (also known as onboard storage) and NAS. Today, users can expect support for multiple languages as well as remote internet access and viewing apps for smartphones for easy access to video from any location.

#### 13.1.1.2 Network video recorder

An NVR is a proprietary hardware box with preinstalled video management software. Usually, users view and manage video through a client software, but many NVR and VMS solutions also offer web interfaces and apps for smartphones and tablets.

The most obvious difference between an NVR platform and a PC server-based solution is that an NVR comes as a hardware box with preinstalled video management functionality (Figure 13.2). In other words, the NVR is a self-contained system that includes the computer, software, storage, and sometimes a multiport Power over Ethernet (PoE) switch in one unit. They are regarded as plug-and-play systems. NVRs are similar to DVRs (digital video recorder) but are used with IP cameras instead of analog cameras. The so-called hybrid DVRs have network inputs as well as analog inputs for recording of both IP and analog cameras.



**Figure 13.2** A network video system that uses a network video recorder.

An NVR is dedicated to its specific tasks of recording, storing, analyzing, and playing back network video. NVRs do not allow any other applications to reside on them. The NVR hardware itself is locked to its application, and the unit can very rarely be altered to accommodate anything outside its original specification.

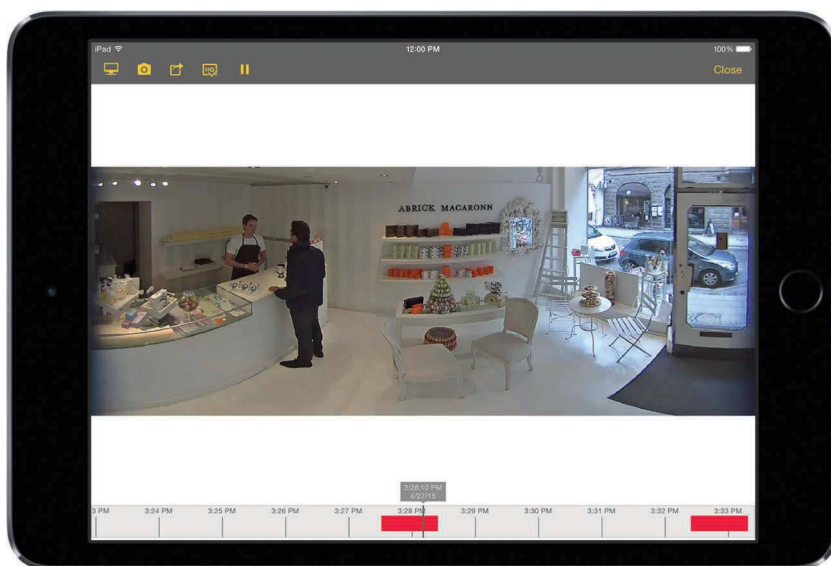
Often, the NVR hardware is proprietary and specifically designed for video management. The operating system can be Windows, UNIX/Linux, or proprietary. An NVR is designed to offer optimal performance for up to a set number of cameras and is normally less scalable than a PC server platform system. This makes the unit more suitable for smaller system configurations where the number of cameras stays within the limits of an NVR's designed capacity. There are low-end and high-end NVR systems available. In comparison with a PC server platform, an NVR is normally easier to install as the video management software comes preinstalled and the NVR sometimes includes built-in PoE ports for quick connection of a few IP cameras.

### 13.1.2 Edge-based video management

Edge-based recording means that no centralized software is needed for the video recording. Instead, the edge device, that is, the camera or video encoder, records to either an SD card or a NAS device. Simply put, the camera also acts as a recorder that saves the video to an SD card or a NAS on the network. This type of system is cost effective and scalable for systems with only a few cameras per site. All recordings can be reached through a smartphone or other connected device.

To make it as user friendly as possible, the process of installing and configuring the system is optimized in a few steps. When finished with the configuration process, the client does need to be opened again. To view live or recorded video and to export video clips, users can use either the PC client or a smartphone or tablet app. The software is often free, but limited to a single camera brand. Its main purpose is to provide easy and reliable video management, with priority on recording video, for businesses with little time or budget for video surveillance. This type of solution is best for small installations of 1–10 cameras, and the sweet spot is somewhere in the middle of that range (Figure 13.3).

In edge-based video management systems, the recorded video is mostly stored on the edge. Many modern network cameras have SD card slots for onboard storage, but even a camera without an SD card slot may support edge storage. In these cases, recordings are typically stored on a NAS



**Figure 13.3** Nowadays, edge-based video surveillance provides a new alternative for small sites, which is not dependent on a central recorder. Edge-based systems often come with user-friendly apps for tablets and smartphones. Here is an example of a viewing app for iPad®.

device placed close to the camera. With edge-based video management, the video analytics is also based on the built-in intelligence of the camera. To view live and recorded video, the user signs into an application on their computer, smartphone, or tablet. There is no need for extra servers, as the camera itself works as a server and keeps track of the recordings, schedules, and events.

Users can also choose if the cameras should record video continuously, to a schedule, or when motion is detected. Through basic functionalities such as support for split-screen viewing, multi-view streaming from multi-megapixel cameras, HDTV resolution, pan, tilt, and zoom (PTZ) control, audio, multiple languages, and search filters, client-camera software meets most of the video management needs of its typical user.

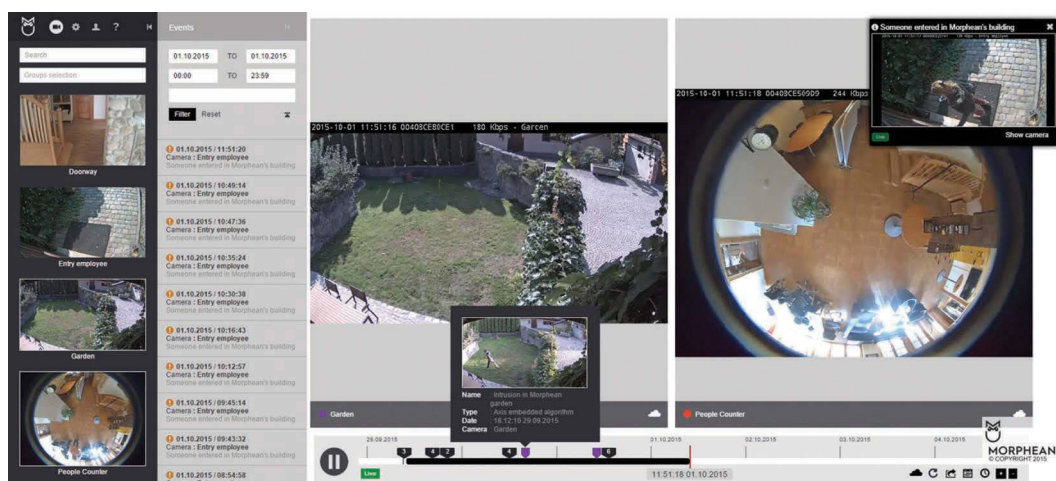
### 13.1.3 Cloud-based video management

Hosted video is a cloud-based video management service that provides remote recording over the internet. Users can view and manage the video system from any device that has a web browser and internet access. Most service providers of cloud video management include apps for smartphones and tablets in their offering. Cloud VMS software is installed on a web server and is connected to one or more recording servers. The web server can allow users from anywhere in the world to access the software and the network video products it manages. All they need is a connection to the internet; a computer, smartphone, or tablet; and a standard web browser (Figure 13.4).

With cloud video management services, also known as hosted video or Video Surveillance as a Service (VSaaS), users do not have to worry about installation or maintenance of servers, clients, or storage devices. The service providers own the storage servers and are responsible for maintenance, and often they have professional operators who can respond to alarms and playback recordings in case of an event. Sometimes, both video management and the actual cameras are included in the monthly subscription fee. However, potential buyers of hosted video do need to make sure that they have a good internet connection, as the transferring of video files consumes a lot of bandwidth.

To compensate for limited bandwidth, you can use complementary technologies. SD cards and a local NAS can store high-resolution, high-frame-rate recordings, while uploading low-resolution video to the cloud. This provides for a very flexible solution.

Hosted video is a good introduction to video management at a relatively low cost and limited up-front investment. It is also a practical solution when a system consists of many locations (sites) with only a few cameras at each site.



**Figure 13.4** Cloud video management software enables users to view cameras and perform various operations with the use of a web browser from any computer, tablet, or smartphone with internet access. (Image courtesy of Morphean, Granges-Paccot, Switzerland.)

One of the major benefits of cloud software is that users only need to pay for the services and the storage they use. The service provider owns all the hardware and is responsible for its maintenance. A good interface should work and look good in most web browsers, no matter the type of device and operating system. Often, the interface is less daunting than client-server software and very easy to use. With the proper safeguards, such as password protection and IP address filtering, cloud services allow for secure online management of video from anywhere in the world.

For more in-depth information about cloud surveillance and its benefits and challenges, see Chapter 14.

## 13.2 OTHER ASPECTS OF VIDEO MANAGEMENT ARCHITECTURE

There are thousands of different video management solutions, from simple onboard recording to fully fledged systems for thousands of cameras integrated with other systems. When deciding upon the best solution, it is also important to consider other aspects of the video management architecture. Some of these aspects are the licensing of the system, how open it is to different types of camera vendors, and which application programming interfaces (APIs) are used for integrating the system, scalability, and support for viewing video on mobile platforms. These aspects are discussed as follows.

### 13.2.1 Open vs. vendor-specific software

Nearly all network camera and video encoder vendors provide their own video management software or NVRs. Some have only very basic recording and viewing functions with limited scalability and may even be given away free to boost sales of network cameras. Others provide more fully fledged systems. As mentioned before, many of these systems only support the network video devices of a specific vendor, which limits choice and flexibility. However, one benefit may be that network cameras and encoders can integrate with better optimization because all the built-in features can be used.

Sometimes, these systems only support the vendors' own devices, but nowadays many support other cameras with open application programming protocols such as ONVIF. Open systems that support multiple brands of network video devices also exist, but they are generally developed by independent companies. Hundreds of versions of open video management software have been made available over the past decade. Although the basic functions are similar, some software is more scalable and includes more advanced functionalities than other software. Because the licensing and maintenance models can differ, it pays to calculate both current and future licenses into the investment cost.

Open systems that support network video devices from many different manufacturers often provide the highest flexibility for the user.

### 13.2.2 Protocols and application programming interfaces

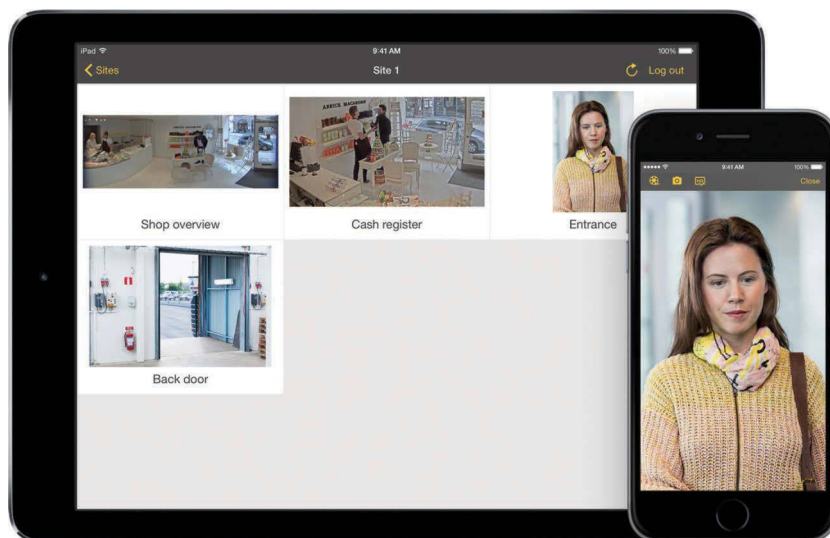
To integrate different network video devices into a video management platform, a communications protocol or an API must be implemented into the video management platform. Most camera vendors use different proprietary protocols. For a few years now, standardized protocols, such as ONVIF, have been developed and have now been adopted by many vendors. While the ONVIF protocol has the benefit of being standardized, it does not always support all advanced features of some cameras. So using the proprietary protocols of a network video manufacturer can provide for additional functionality and tighter integration.

### 13.2.3 Apps for smartphones and tablets

As mentioned in the previous sections, many video management software vendors complement their offerings with apps for smartphones and tablets (Figure 13.5).

The level of functionality varies, from viewing of live video only to some level of video management. Users are now starting to expect a higher level of intelligence and control, as well as more integration with other systems, such as the ability to greet visitors and open doors from a smartphone.





**Figure 13.5** Viewing apps for tablets and smartphones provide mobile access to live and recorded video. This example is of a viewing app for iPad® and iPhone®.

### 13.2.4 Scalability of video management software

The scalability of most video management software, in terms of the number of cameras and frames per second that it supports, is in most cases limited by the hardware and bandwidth capacity rather than by the software itself. Storing video files puts new strains on the storage hardware as it may be required to operate on a continual basis, as opposed to only during normal business hours. In addition, video generates large amounts of data, placing high demands on the storage solution.

The basic recording of video and audio is not particularly CPU intensive. More processing power is needed when frames require decoding, as is the case with functionalities such as VMD and other server-based video analytics. Live viewing capacity is limited by the capacity of the video decoder and graphics card.

As a system grows and eventually exceeds the capacity of a single server, it is in many cases possible to add new servers to the system or pay a higher fee to the cloud service provider. Ideally, it should be possible to add storage seamlessly. For professional and scalable systems, factors such as redundancy, failover, and no single point of failure become very critical. Redundancy in a storage system allows for saving video, or other data, simultaneously to more than one location. This provides a backup for recovering video if a portion of the storage system becomes unreadable. There are a number of options for providing this added storage layer in a network video system, including a redundant array of independent disks, data replication, server clustering, and multiple video recipients. See Chapter 12 for more on servers and storage.

### 13.2.5 Licensing of video management software

Most VMS applications are licensed products. Licensing policies vary but in most cases, they involve one license per camera or video source in the system as well as a base fee for the software itself. Some VMS software only requires one license per encoder, regardless of whether it is a multichannel or single-channel encoder, while others require one license per channel. Sometimes, the license is a software key tied to the MAC address of a camera or a server CPU. Typically, there are also maintenance fees associated with technical support and the right to download and install new versions of the software. In the case of cloud video management, all services are included in the monthly fee to the service provider.



## 13.3 SYSTEM FEATURES

A video management system can include many different features. Some of the more common ones are:

- Recording of video
- Recording of audio
- Simultaneous viewing of video from multiple cameras
- Event management functions
- Camera administration and management
- Search options and playback
- User access control and activity (audit) logging
- Video analytics functionality

### 13.3.1 Recording

The primary function of a video management system is to record video and audio. Recording functionalities include setting up recording rules and intelligent ways of searching recordings and exporting them to other systems. The following subsections discuss the various recording aspects of a video management system.

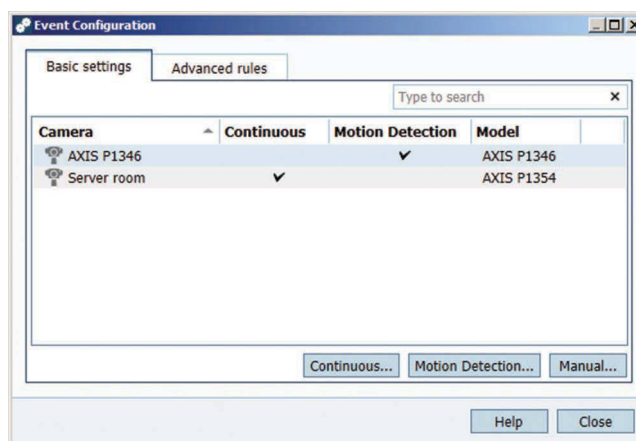
#### 13.3.1.1 Video recording

A typical video management system has three types of video recordings:

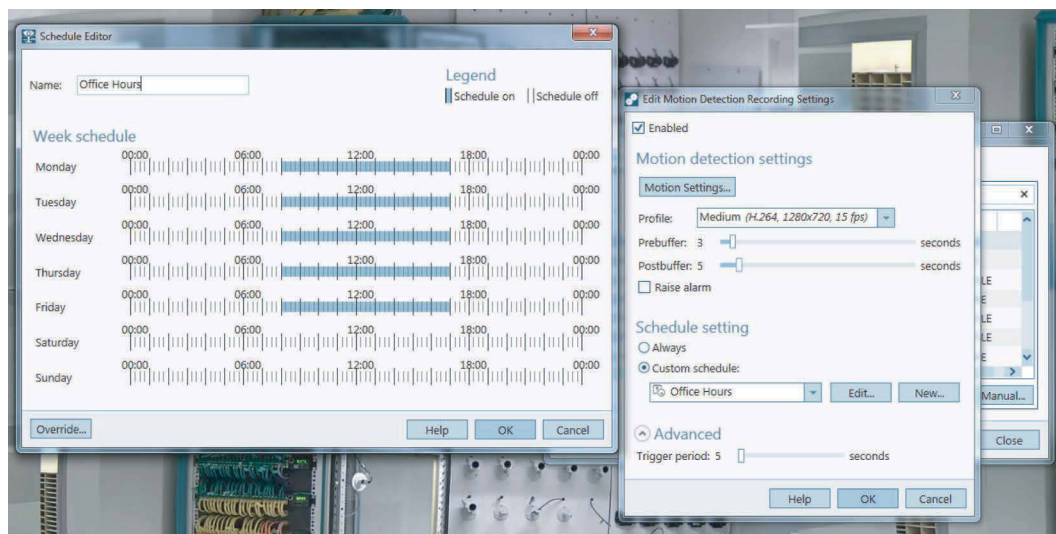
1. Continuous recordings
2. Triggered recordings (by motion or alarm)
3. Scheduled recordings

Because they tend to last for longer periods, continuous recordings normally use more disk space than alarm-triggered recordings. An alarm-triggered recording can be activated by, for example, VMD or by an external input on a camera's input port. Scheduled recordings can combine time-tables for both continuous and triggered recording instructions (Figures 13.6 and 13.7).

After selecting the recording method, the quality of the recordings is determined by selecting the video format (e.g., Motion JPEG or H.264), resolution, frame rate, and degree of image compression. These parameters affect how much bandwidth and storage space is needed.



**Figure 13.6** An example of an interface for editing recording functions.



**Figure 13.7** Scheduled recording settings with a combination of continuous and alarm or motion recordings applied.

Each recording mode can have its own frame rate (frames per second). Because of previous limitations of analog recording technologies (DVRs), the specification and requirements of many system designs are limited to lower resolutions such as VGA or 720p, and frame rates of 7.5 frames per second (fps) or lower are common. With today's network cameras and video compression technologies, higher resolutions and resolutions are feasible. Video streams at 5 megapixels or 4K, with frame rates of 30 fps or even higher, are not unreasonable.

Storage is still a relatively high expense in the video surveillance system, so making the right requirements on the system continues to be crucial. In a well-designed system, the network camera or encoder requests the appropriate frame rate, which means that only the required frame rate is sent over the network.

A VMS application usually has a background service that automatically starts running upon system start-up. When the background service is running, recording will continue even after a user has logged off.

### 13.3.1.2 Audio recording

Audio can be recorded using the built-in functionality in a network video product or through the video management software. When it is important that audio and video are synchronized, the audio and video packets must be time-stamped. Because a network camera might not always support time-stamping of Motion JPEG video packets, using H.264 is recommended. H.264 video streams are normally sent using Real-Time Transport Protocol (RTP), which time-stamps the video packets. For details on audio compression standards, see Chapter 7. Some areas have restrictions on audio recordings. For more information about legal considerations, see Chapter 17.

### 13.3.1.3 Recording and storage

Most video management software uses the standard Windows file system for storage, so any system drive or network-attached drive can be used for storing video. Usually, a separate file or database keeps an index of the available video. The advantages of using a database for storing all settings and recording metadata, and using a file system, include:

- Ability to manage shared access and ensure data integrity
- Ability to efficiently search for recordings
- Ability to enable direct file access and record directly to disk

Video management software can support more than one level of storage. That is, a primary hard drive records video, but local disks, network-attached drives, or remote hard drives archive the files. Users may be able to specify for how long files should remain on the primary hard drive before they are automatically deleted or moved to the archive drive. Users also may be able to lock video files and protect them from automatic deletion.

#### 13.3.1.4 Search options

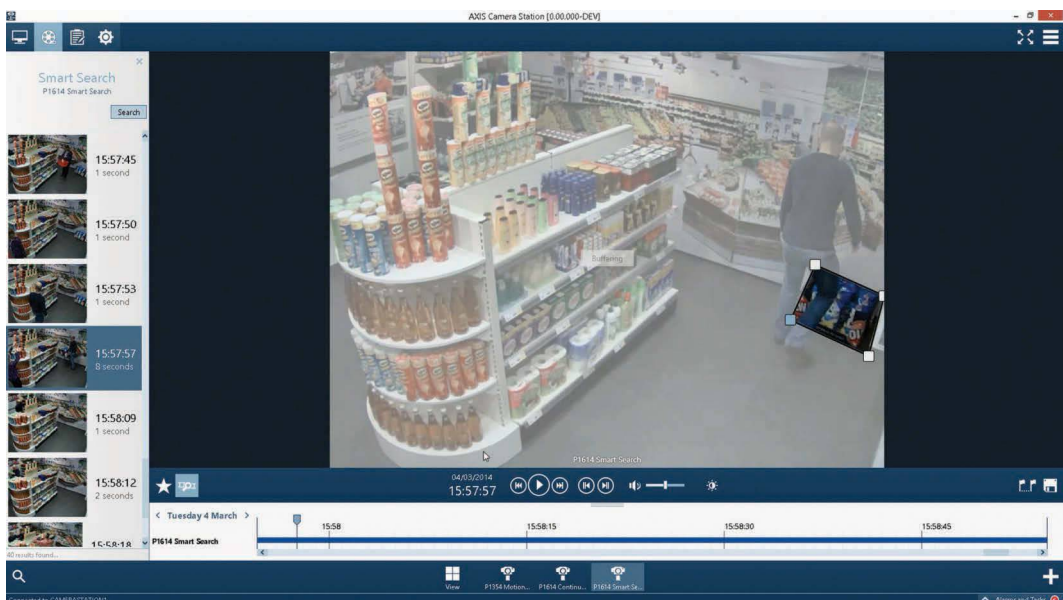
Searching for recorded video is an important feature in a video management system. There are different ways of finding the footage of interest, including the following:

- Scrubbing through footage
- Selecting dates and times
- Activity-based or event-based searches

Scrubbing means that the user manually goes through the recorded video to find certain activity in the scene. Because there are often many hours of footage recorded, scrubbing can be an ineffective method. If the user knows the date and time an event occurred, they can enter that time and replay video sequences from one or more cameras. If the user only has evidence that something has happened, but not when, they can do a motion-based search (Figure 13.8). The video management system finds and replays video of movements that occur within a defined area of interest. The system ignores all events outside that area. This is the most advanced type of search and is also known as advanced search, quick search, or smart search. Naturally, it is easy to find and replay video associated with a certain type of alarm or event trigger.

#### 13.3.1.5 Exporting files

Many VMS applications record video using proprietary file formats, but still allow users to export recorded video to standard file formats such as mp4, Audio Video Interleave, and Advanced Streaming Format. In many cases, the proprietary format can be exported along with a proprietary player. The advantages of using the native recording format are additional security, data integrity, and advanced playback features. The proprietary format can be encrypted and password protected.



**Figure 13.8** In video management systems, events in a defined area can easily be found through motion-based smart search.

Because the proprietary format is often more difficult to edit, it can be used to preserve the chain of evidence. A proprietary player may also include other features, such as multicamera playback.

## 13.3.2 Viewing

Another key function of a video management system is the ability to view live and recorded video in an efficient and user-friendly way. Most video management software lets multiple users view several different cameras at the same time and allows recordings to occur simultaneously. Additional features include multimonitor viewing and mapping, the latter meaning that camera icons can be overlaid on a map of the building or area, representing the true location of each camera. The following subsections discuss common viewing functionalities in video management systems.

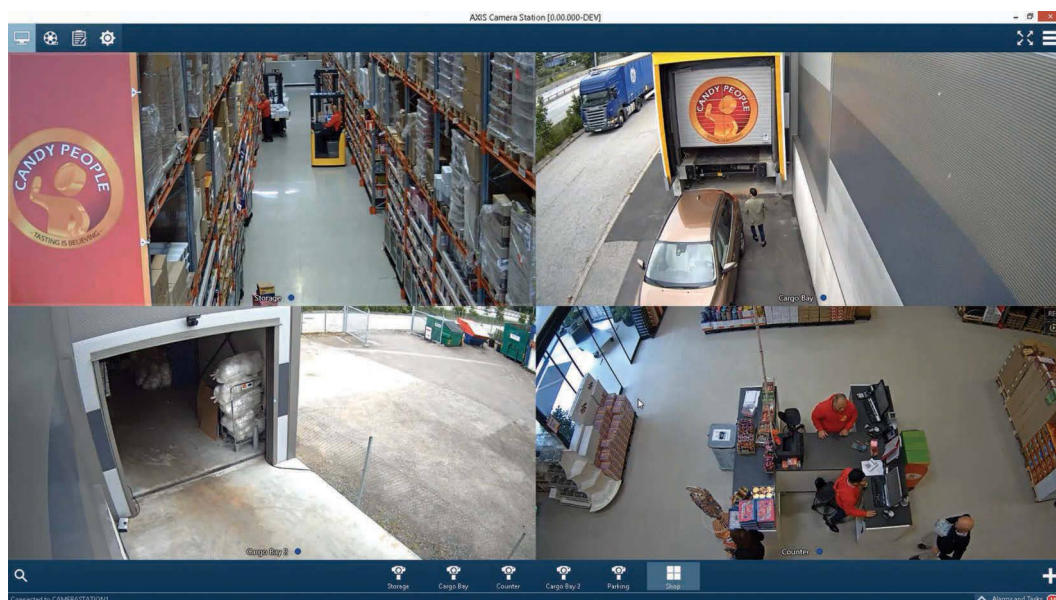
### 13.3.2.1 Live viewing

Many VMS applications provide users with the option of viewing images in different ways: split views, single-camera pop-ups, full screen, or camera sequences. In a split view, between 4 and 64 cameras are shown on the screen. A pop-up can also appear when a camera detects motion. Figure 13.9 shows an example of a 2×2 split view. In sequence mode, live views from different cameras will be displayed, one after the other, in a user-defined order. Users can choose to switch between cameras, view areas, and split views. They can also set a time for each view in the sequence, that is, the time to elapse before switching to the next view.

When using a PTZ camera, video management software may allow PTZ control through a number of input devices, including the following:

- Joysticks
- Control boards, keyboards, and keypads
- Computer mouse

Many camera manufacturers supply joysticks and control boards especially developed for video surveillance (see Figure 13.10), but even a gaming controller could be used, as long as it is compatible with the VMS and the camera's PTZ protocol. Users can program control boards, keyboards, and keypads with shortcuts, or hotkeys, to move quickly between workspaces, camera views, and



**Figure 13.9** An example of a split view.





**Figure 13.10** A control board consisting of a keypad, jog dial, and joystick.

PTZ positions. They can also use a mouse to click on the image, move the camera, and use the scroll wheel to zoom in.

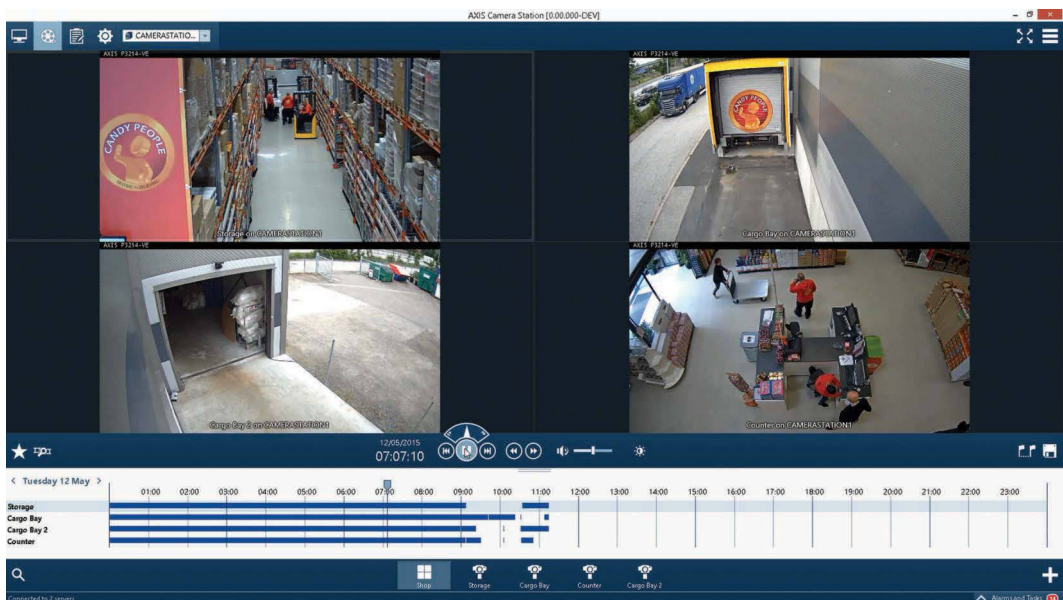
If a camera is equipped with audio capability, the VMS can also allow audio controls through the user interface.

### 13.3.2.2 Viewing of recordings

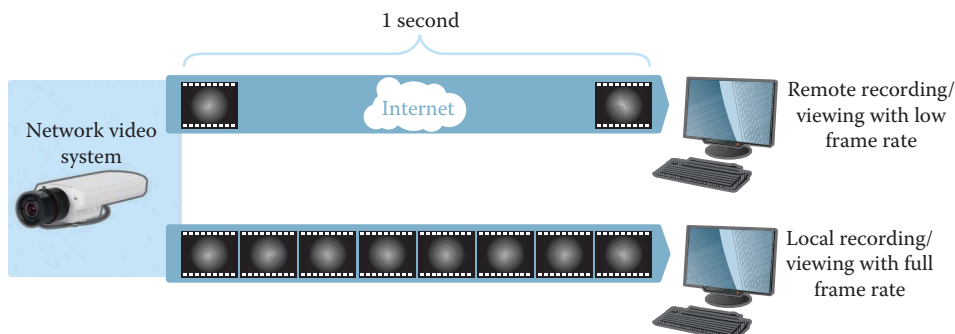
Video management software usually includes multicamera playback, which lets users view simultaneous recordings from different cameras (Figure 13.11). This makes it easier to get a comprehensive picture of events, which both moves investigations forward and helps build cases in court.

### 13.3.2.3 Multistreaming

Viewing or recording at full frame rate on all cameras at all times is more than what is required for most applications. In normal operation, frame rates can be set lower, say one to four frames per second, to limit bandwidth and storage requirements. The recording frame rate can be set to increase automatically when the camera detects particular types of events, such as motion in a monitored area or activation of an external sensor. It is also possible to send multiple video streams in parallel with different frame rates, codecs, compression, and resolution to different recipients (Figure 13.12), taking into consideration the available bandwidth and the performance of the devices used for recording and viewing. This ability to create, view, send, and record multiple streams with different sets of capabilities and values is called multistreaming.



**Figure 13.11** Split views help the operator get a better overview of when and how a certain incident evolved.



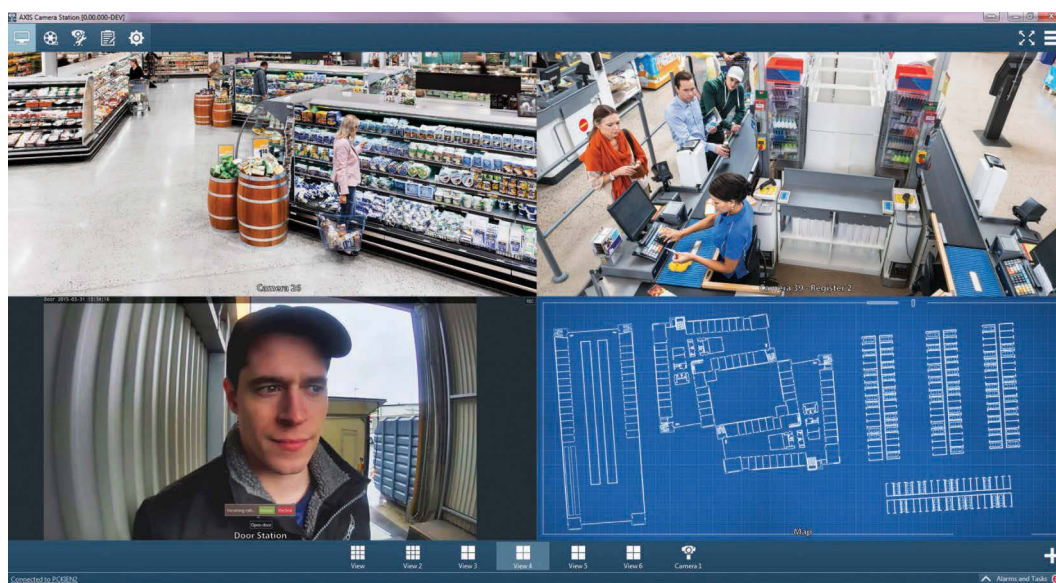
**Figure 13.12** Different frame rate video can be sent to different recipients.

Multistreaming is great when you want flexible recording, viewing, and storage options. For example, you can use a low-quality, low-frame-rate stream for viewing in smartphones and tablets to save bandwidth when accessing video remotely or to offload a software client with poor performance. In an edge-based video management system, each camera can record parallel streams (one low-bandwidth and one high-quality stream) and keep track of its recordings. Only when you request a stream, the camera sends it to your viewing device. Of course, streaming the low-bandwidth video saves data traffic over the mobile network, but you can switch to the high-quality stream whenever you want to. Say that you are watching a low-bandwidth stream until you suddenly spot something interesting. By switching to the high-quality profile, you can watch the event in greater detail. You can also save video and images on your mobile device and share them with the police, your insurance company, or other investigators.

### 13.3.2.4 Mapping functionality

For easy identification and selection of cameras, users can place camera icons on a map of the monitored area. The map is imported into the video management software (Figure 13.13) and can be a photo or a drawing in a standard image format such as jpg, gif, or png.

Some video management systems have an icon library that makes it possible to drag and drop icons onto a map. These icons can represent different types of cameras. By clicking on a camera icon,



**Figure 13.13** Using a mapping functionality, finding the right camera becomes very intuitive.



the user can view the live video stream from that camera. When an alarm occurs, a camera icon may change color to indicate that the alarm comes from that camera. More advanced systems also make it possible to show which area each camera covers.

### 13.3.3 Event management

Video management software and NVRs usually have the ability to receive, process, and associate events from different sources such as PACSs, point-of-sale (POS) terminals, video analytics software, and the network video products themselves. Once an event is triggered, the video management system can register the event, associate it with a video clip from a nearby camera, and alert an operator or investigator through a pop-up window on the monitor or by sending a notification to a smartphone or tablet.

The following subsections provide more details about event and alarm management, VMD, input/output (I/O) ports, and event log files.

#### 13.3.3.1 Edge-based event handling

Network video products with built-in intelligent processing have the huge benefit of enabling efficient use of bandwidth and freeing up network storage space. Unless an event takes place, there is no need for such a network camera or video encoder to send continuous video streams for live monitoring or recording. When an event takes place, the network video product can send notifications, trigger other devices, and activate other responses automatically.

Event management, which includes alarm handling, involves defining an event that activates a network video product to perform certain actions. An action can be scheduled or triggered. See Figure 13.14 for a trigger setup example.

There are many different types of triggers that can activate actions, including the following:

- *Inputs* from external devices can trigger actions. The input devices are connected to the input ports on a network camera or video encoder. Examples of input devices are motion sensors, door switches, and glass-break detectors. See Figure 13.15 for an I/O example.
- *Manual triggers* are used by operators to initiate actions.
- *VMD* can trigger actions when the camera detects movement based on settings made in user-defined motion detection windows.
- *Audio detection* can trigger actions when a camera with built-in audio input detects sound below or above a set threshold value. For more information on audio detection, see Chapter 7.

**Action Rule Setup**

**General**

☒ Enable rule

Name:

**Condition**

Trigger:

Audio Detection

Above alarm level: ☒ Yes ☐ No

Schedule:

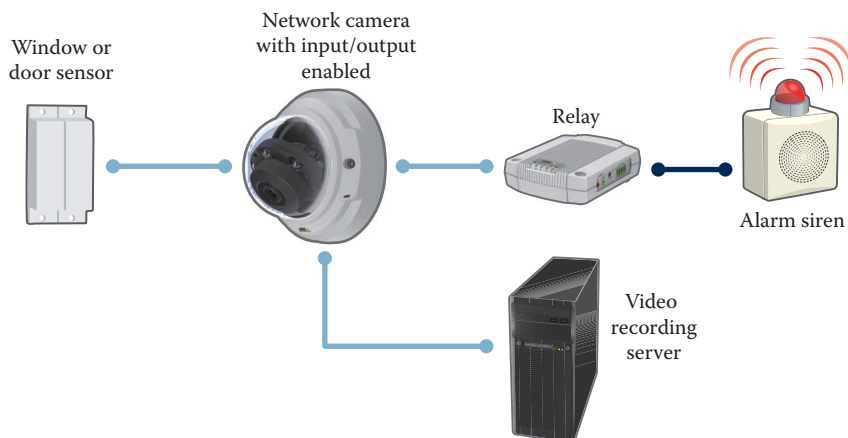
☐ Additional conditions

**Actions**

Type:

- Day/Night Vision Mode
- Overlay Text
- Play Audio Clip
- PTZ Control
- Record Video**
- Send Images
- Send Notification
- Send SNMP Trap
- Send Video Clip

**Figure 13.14** An example of setting up an action rule that triggers an action (e.g. recording of video) when an event occurs (e.g. the camera picks up sound).



**Figure 13.15** A window sensor can connect to the camera's input port and an alarm system or a siren to the output port.

- *Temperatures* can trigger actions when the camera detects that the temperature rises or falls outside the operating range.
- *Camera tampering* can trigger actions when the camera detects that it has been covered, moved, or is no longer in focus. For more information on camera tampering, see Chapter 16.
- *Shock detection* can trigger actions when the camera detects that someone has tilted it or hit it with an object.
- *Storage recording* can trigger actions when the camera starts or stops recording video to edge storage.
- *Storage disruption* can trigger actions when the camera detects that the storage device is unavailable, locked, full, or removed or if there are read/write errors.
- *Enter/exit detectors* can trigger actions when the camera detects that a subject or object enters or exits a user-defined area.
- *Fence detectors* can trigger actions when the camera detects that a subject or object crosses a virtual line.
- *Object removed* can trigger actions when the camera detects that an object disappears from a user-defined area.
- *PTZ error* can trigger actions if the camera's PTZ functionality stops working correctly.
- *Passive infrared detector (PIR) sensors* can trigger an event in the camera if a PIR sensor, built-in or connected to the camera's I/O port, has detected activity.

### 13.3.3.2 Responses

Network video products can respond to events all of the time or at specific times. When a trigger is activated, some of the common responses and actions that can be configured include the following:

- *Recordings* can be used to record images, video, and audio to specified locations, at specified frame rates and compression during the course of an event.
- *Activate outputs* can be used to activate external devices, such as sirens, lights, and relays connected to the output ports on a network camera or video encoder. (see Figure 13.15)
- *Send email notifications* can be used to send emails to users, notifying them that an event has occurred. The email can include an image of the event.
- *Send HTTP or TCP notification* can be used to send files and alerts to a video management system. In turn, this can trigger an action in the video management software such as a recording or activation of another camera or auxiliary device.

- *Send an SMS/text message* can be used to send a text or multimedia message with information about the event.
- *Save to edge storage* can be used to record video, and sometimes audio, to an edge storage device such as an SD card or a NAS.
- *Go to PTZ preset* can be used to get a PTZ camera to move automatically to a specified position (e.g. a gate, fence, or entrance) when an event takes place.
- *On-screen pop-up* can be used to make a live view window pop-up on the operator's monitor when an event takes place.
- *On-screen instructions* can be used to give the operator a procedure to follow when an event takes place.

Through prealarm and postalarm buffers, recordings can include video of not just the event but also a set amount of time before the trigger and after the event ends. This can give a more complete picture of the event.

### 13.3.3.3 Video motion detection

Video motion detection (VMD) is a way of defining activity in a scene by analyzing image data and differences in a series of images. VMD and other video analytics allow video surveillance to be event driven rather than operator driven. In many cases, it can shorten the response time and the intervention efficiency, as well as reducing the time security staff must spend observing live video and patrolling the area.

As mentioned before, many cameras have built-in VMD. This means that a network camera will send video to the software application for analysis. For cameras that do not have built-in VMD, the video management software can provide this feature instead. Using VMD makes it easier to prioritize recordings, decrease the amount of recorded video, and search for events. Although the principle of VMD is more or less the same whether it is built into the camera or is part of the video management software, the two solutions affect the infrastructure and usage of the video management system in different ways. The following paragraphs discuss some of these similarities and differences.

#### 13.3.3.3.1 VMD in video management software

VMD can detect motion in any area of an image. Usually, users define specific areas of interest in which the software should react to movement. They may also be able to set different sensitivity levels, adjusting for which object and movement size is typical for the monitored area (Figure 13.16). Upon detection of motion, the software can trigger an external device to perform an action (such as open or close a door or turn a light on or off), start recording video from selected cameras, and send emails. VMD can also be used to trigger actions if motion stops. This is useful in scenarios where a specific type of motion is expected, for example, in factories where robots and machines have very predictable movements.

Remember that running VMD in the video management software is a CPU-intensive process and if used on many video channels it puts a lot of strain on the system.

#### 13.3.3.3.2 VMD at the edge

Using the built-in VMD (also known as embedded VMD) in a network camera or video encoder offers substantial advantages over using the video management software's VMD. Because the network camera or video encoder performs the motion detection process itself, it frees up CPU power that can be used for other processes and recording devices in the system. Edge VMD helps make video surveillance more event-driven and cuts the demands on bandwidth and storage because no video, or only low-frame-rate video, streams to the operator or recording devices while there is no activity in the scene.

The built-in VMD in network cameras (Figure 13.17) or video encoders is very similar to the VMD functionality found in video management software. Users can configure motion in certain areas while ignoring motion in others.

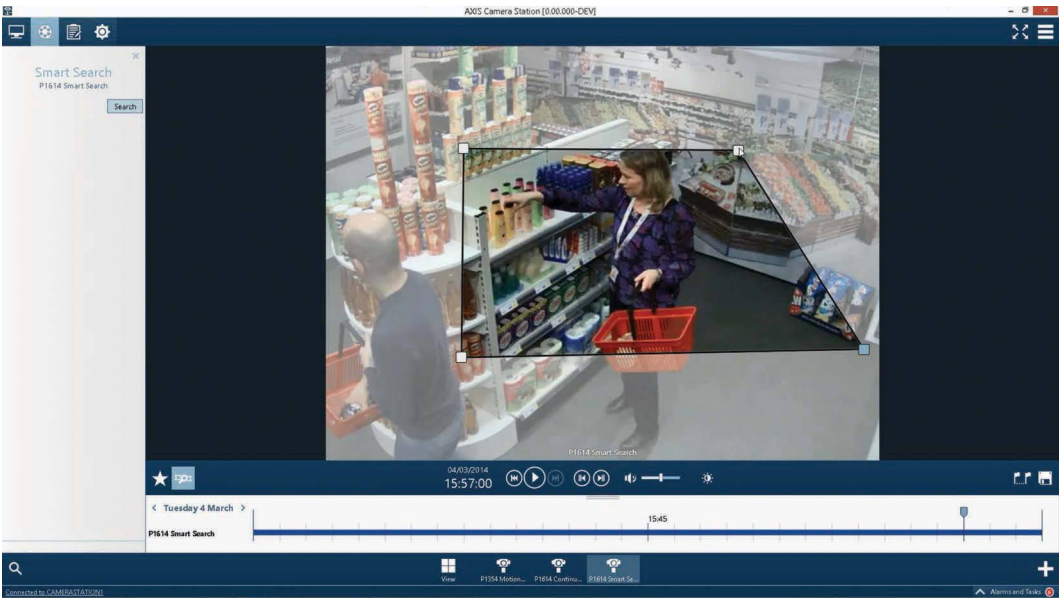


Figure 13.16 Setting video motion detection in video management software.

### AXIS Video Motion Detection 3

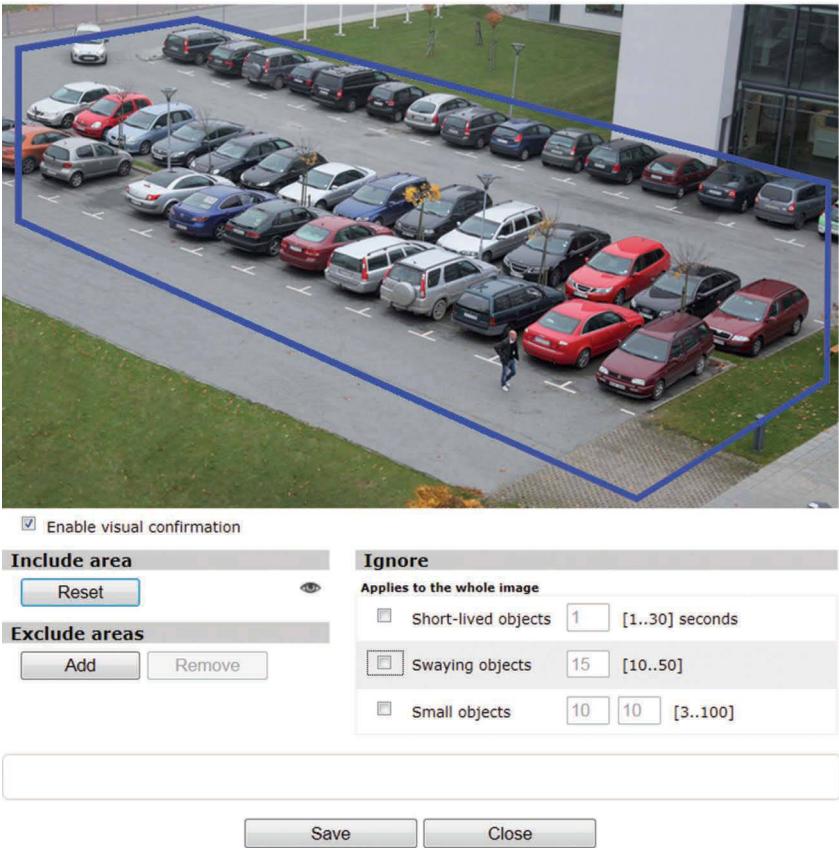


Figure 13.17 Configuring video motion detection in a network camera.

VMD data that provide information about, for example, the level of activity or size of a moving object can be included in a video stream to simplify searches in the recorded material. For more details about VMD, see Chapter 16.

VMD is only one type of video intelligence. Other common analytics include camera tampering and crossline detection. For more examples, see Chapter 16.

13.3.3.4 I/O ports

Many network cameras and network video encoders have integrated I/O ports. External devices can connect to these ports so that they can communicate with and be managed by the system. For example, a camera that receives an input from an external alarm sensor can be instructed to only send video when triggered by the sensor.

The range of devices that can connect to a network video product’s I/O ports is almost endless (Tables 13.1 and 13.2). The basic rule is that any device that can toggle between an open and closed circuit can connect to a network camera or a video encoder. The main function of a network video product’s output port is to trigger external devices, either automatically or by remote control from an operator or a software application.

13.3.3.5 Event log files

Video management software can provide an event log (Figure 13.18) that includes a list of camera and server events based on date, time, type, and source of the events. Users can sort or search for specific events with or without filters, for example, a list of I/O activities that occurred between two specific times or when motion was detected and where.

**Table 13.1** Examples of devices that can connect to an input port

Device type	Description	Usage
Door contact	A magnetic switch that detects the opening of doors or windows	When the circuit opens (the door opens), the camera performs an action, for example, sending full-motion video and notifications.
PIR	A sensor that detects motion based on heat emission	When the sensor detects motion, it opens the circuit and the camera performs an action, for example, sending full-motion video and notifications.
Glass-break detector	An active sensor that measures air pressure in a room and detects sudden pressure drops	When the sensor detects an air pressure drop, it opens the circuit and the camera performs an action, for example, sending full-motion video and notifications.

**Table 13.2** Examples of devices that connect to an output port

Device type	Description	Usage
Door relay	A relay (solenoid) that controls the opening and closing of door locks	A remote operator can (over a network) lock and unlock a door.
Siren	Alarm siren configured to sound when alarm is detected	When motion is detected (through VMD or a digital input), the camera can activate the siren.
Alarm/intrusion system	An alarm security system that continuously monitors a normally closed or open alarm circuit	The camera can act as an integrated part of the alarm system that serves as a sensor, enhancing the alarm system with event-triggered video transfers.

Time	Category	Message
2015-11-13 16:12:01	Info	Recording stopped on AXIS P3905-R
2015-11-13 16:11:56	Info	Motion detected on AXIS P3905-R
2015-11-13 16:11:42	Info	Recording stopped on AXIS P1427-LE
2015-11-13 16:11:41	Info	Recording stopped on AXIS P1428-E
2015-11-13 16:11:37	Info	Motion detected on AXIS P1427-LE
2015-11-13 16:11:35	Info	Motion detected on AXIS P1428-E
2015-11-13 16:11:26	Info	Recording started on AXIS P1427-LE
2015-11-13 16:11:26	Info	Motion detected on AXIS P1427-LE
2015-11-13 16:11:26	Info	Recording started on AXIS P1428-E
2015-11-13 16:11:26	Info	Motion detected on AXIS P1428-E
2015-11-13 16:11:23	Info	Recording started on AXIS P3905-R
2015-11-13 16:11:23	Info	Motion detected on AXIS P3905-R
2015-11-13 16:10:59	Info	Recording started on AXIS P1365
2015-11-13 16:10:59	Info	Recording started on AXIS M4054
2015-11-13 16:10:39	Info	Recording started on AXIS P1405-LE
2015-11-13 16:03:06	Info	Server PCREMY11 started
2015-11-13 16:03:47	Info	Server PCREMY11 started

**Figure 13.18** An example of an event log.

### 13.3.4 Administration and management features

A video management software should be able to handle the administration of cameras and video encoders. This includes installation, firmware upgrades, security, audit log, and parameter configurations. If the system includes other surveillance-related devices, such as cash registers, door controllers, and door stations, it simplifies things if the software can be used to handle those as well. The larger the surveillance system, the more important it is to be able to efficiently manage the networked devices.

#### 13.3.4.1 Managing cameras

All video management software provides the ability to add and configure basic camera settings, frame rate, resolution, and compression format, but some also include more advanced functionalities such as camera discovery and complete device management.

Software that helps simplify the management of networked cameras, video encoders, and other devices in a system often provides the following functionalities:

- Locating and showing connection status of video devices on the network
- Setting IP addresses
- Configuring single or multiple devices
- Managing firmware upgrades of multiple devices
- Managing user access rights

Video management software can provide a configuration sheet (Figure 13.19) that gives users an overview of all camera and recording configurations.

#### 13.3.4.2 Time synchronization

Conducting an investigation using multiple cameras or integrating different systems becomes easier if all networked devices have the same time. The most common way to achieve this is to use a network time protocol (NTP) server to synchronize all devices in a network. Most networked devices support NTP (Figure 13.20).



## Server Configuration Sheet for AXIS Camera Station

### General Configuration Information

Program Version	0.00.000-DEV
Protocol Version	4.3.0
Process	64bit
Application Culture	en-US
DirectX Version	4.09.00.0904
.NET CLR Version	4.6 or later (release 393297, CLR 4.0.30319.42000)
Entry Assembly	AcsService
Server Name	PCREMYJ1
Server GUID	198056c-a767-4710-817b-db8e096df235
Generated UTC	2015-11-13 15:15:23
Install Path	C:\Program Files\Axis Communications\AXIS Camera Station
Is Version 2 Service Running	No
OLE/DB Exchange	91835810327701627
Operating System	Microsoft Windows 7 Enterprise Service Pack 1
OS Culture	en-US
OS Version	6.1.7601.65536
Is Part Of Domain	Yes
Database size	11.9 MB

### Camera Settings

Camera Name	ID	Type	Firmware Version	Address	HTTP Port	Hostname	Serial Number	Video Port	Device ID	GUID	Is Enabled	Disconnects Since Server Start	Audio Source Device ID	High Profile	Medium Profile	Low Profile	External Live View Profile	External Recording Profile	Vmd Version	Supports Metadata	Include Analytics Data	Installed Applications
AXIS M5014	135	AXIS M5014	5.50.3.1	172.25.183.196	80		ACCC8E1A202E	1	133	d34a05a1-f0e7-4911-986d-7b34c8e64ef1	Yes	0	Internal	Live View Audio=No, Record Audio=No FPS=25, Format=H.264, Resolution=1280x720, Compression=30	Live View Audio=No, Record Audio=No FPS=15, Format=H.264, Resolution=1280x720, Compression=30	Live View Audio=No, Record Audio=No FPS=5, Format=H.264, Resolution=640x360, Compression=30	Live View Audio=No, Record Audio=No FPS=15, Format=H.264, Resolution=1280x720, Compression=30	Live View Audio=No, Record Audio=No FPS=15, Format=H.264, Resolution=1280x720, Compression=30	Motion Detection	No	No	
AXIS P1365	84	AXIS P1365	5.75.3.1	172.25.183.101	80		ACCC8E0203FE	1	82	0c0b45aa-fca0-4c9c-943d-49e5743bf4c4	Yes	0	Internal	Live View Audio=No, Record Audio=No FPS=30, Format=H.264, Resolution=1920x1080, Compression=30	Live View Audio=No, Record Audio=No FPS=15, Format=H.264, Resolution=1280x720, Compression=30	Live View Audio=No, Record Audio=No FPS=5, Format=H.264, Resolution=640x360, Compression=30	Live View Audio=No, Record Audio=No FPS=15, Format=H.264, Resolution=1920x1080, Compression=30	Live View Audio=No, Record Audio=No FPS=15, Format=H.264, Resolution=1280x720, Compression=30	Video Motion Detection 2.1	Yes	No	Metadata, VideoMotionDetection
AXIS P1405-LE	78	AXIS P1405-LE	5.80.1.1	172.25.183.96	80		ACCC8E0CB716	1	76	56328690-9a07-4523-9a2f-951cb02160f2	Yes	0	None	Live View Audio=No, Record Audio=No FPS=25, Format=H.264, Resolution=1920x1080, Compression=30	Live View Audio=No, Record Audio=No FPS=15, Format=H.264, Resolution=1280x720, Compression=30	Live View Audio=No, Record Audio=No FPS=5, Format=H.264, Resolution=640x360, Compression=30	Live View Audio=No, Record Audio=No FPS=15, Format=H.264, Resolution=1920x1080, Compression=30	Live View Audio=No, Record Audio=No FPS=15, Format=H.264, Resolution=1280x720, Compression=30	Video Motion Detection 2.1	Yes	No	Metadata, VMD3, VideoMotionDetection
AXIS P1425-	69	AXIS P1425-	5.80.1.1	172.25.183.98	80		ACCC8E0C942F	1	67	cca9f377-b855-43fb-b866	Yes	0	None	Live View Audio=No, Record Audio=No FPS=25, Format=H.264	Live View Audio=No, Record Audio=No FPS=15, Format=H.264	Live View Audio=No, Record Audio=No FPS=5, Format=H.264	Live View Audio=No, Record Audio=No FPS=15, Format=H.264	Live View Audio=No, Record Audio=No FPS=15, Format=H.264	Video Motion Detection	Yes	No	Metadata, VMD3, VideoMotionDetection

Figure 13.19 An example of a configuration sheet.

**Figure 13.20** Most network cameras and video encoders have support for network time protocol.

### 13.3.4.3 Security

An important part of video management is security. A network video product or video management software should have options for the following settings:

- Authorized users
- Passwords
- Multiple user access levels
- User-differentiated access to specific devices

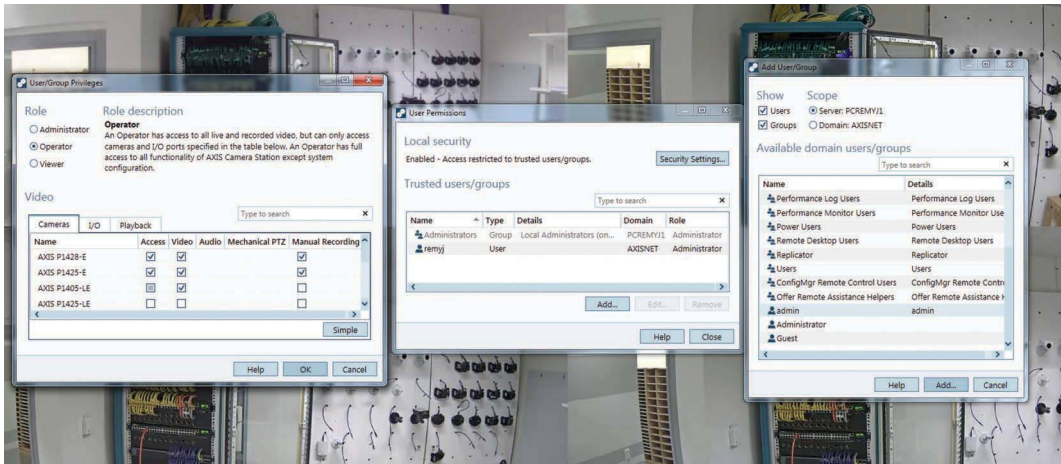
Through user-differentiated access, the system administrator can specify exactly which users and groups have access to each networked device. The access levels (Figure 13.21) define which information users can see and which changes they can make. The following user access levels are common in a network video context:

- Administrators have access to all functionalities.
- Operators have access to all functionalities except for certain configuration pages.
- Viewers only have access to live video.

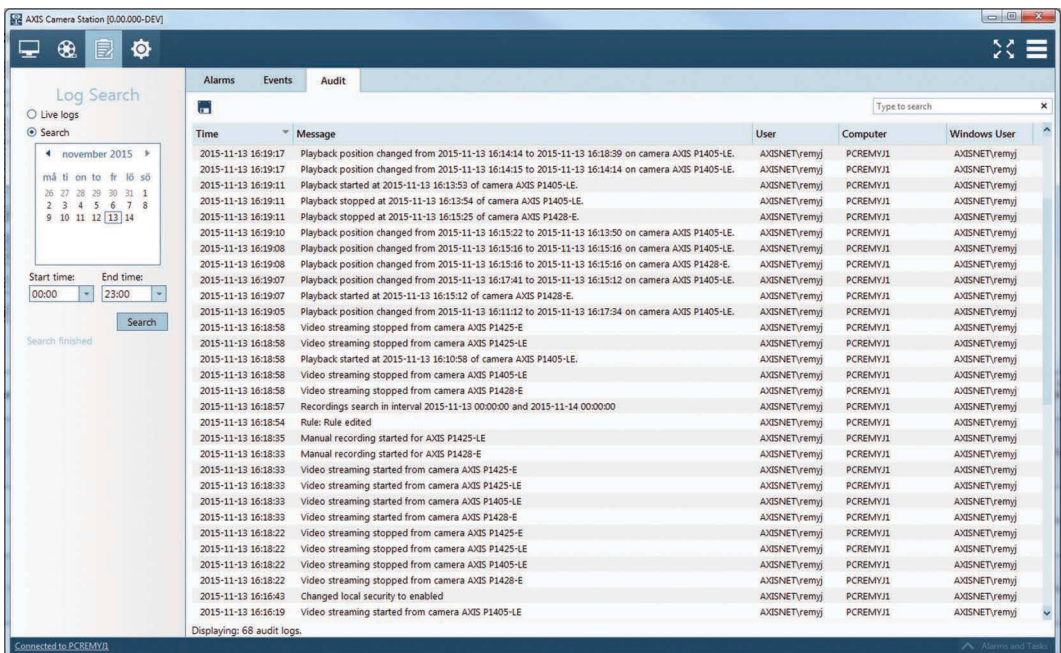
Many VMS programs can inherit a Windows user database (local or LDAP/domain). This feature eliminates the need to set up and maintain a separate database of users.

### 13.3.4.4 Audit log files

The VMS's audit log shows a list of user actions based on the user, time, type of activity, and video device (Figure 13.22). The audit log file is an essential function that provides proof of who used the system and when, what happened, and which actions the user took.



**Figure 13.21** An example of interface where user access can be set.



**Figure 13.22** An audit log lets users generate a list of user actions. Fields such as the user, time, type of activity, and camera can be filtered or sorted.

## 13.4 INTEGRATED SYSTEMS

Video management systems that are based on a network video platform can be easily integrated into other IP-based systems such as POS, physical access control, building management, and industrial control systems. When video is integrated, information from other systems can be used to trigger functions such as event-based recordings in the network video system, and vice versa. In addition, having a common interface for managing different systems makes it easier for users to do their jobs.

### 13.4.1 Application programming interface

An application programming interface (API) enables the development of customized applications. A video management system must have an API to integrate the video system into other systems.

Once implemented, the subsystems can communicate with each other and perform actions such as starting recordings, sending alarms, opening doors, activating microphones and speakers, and accessing live and recorded video.

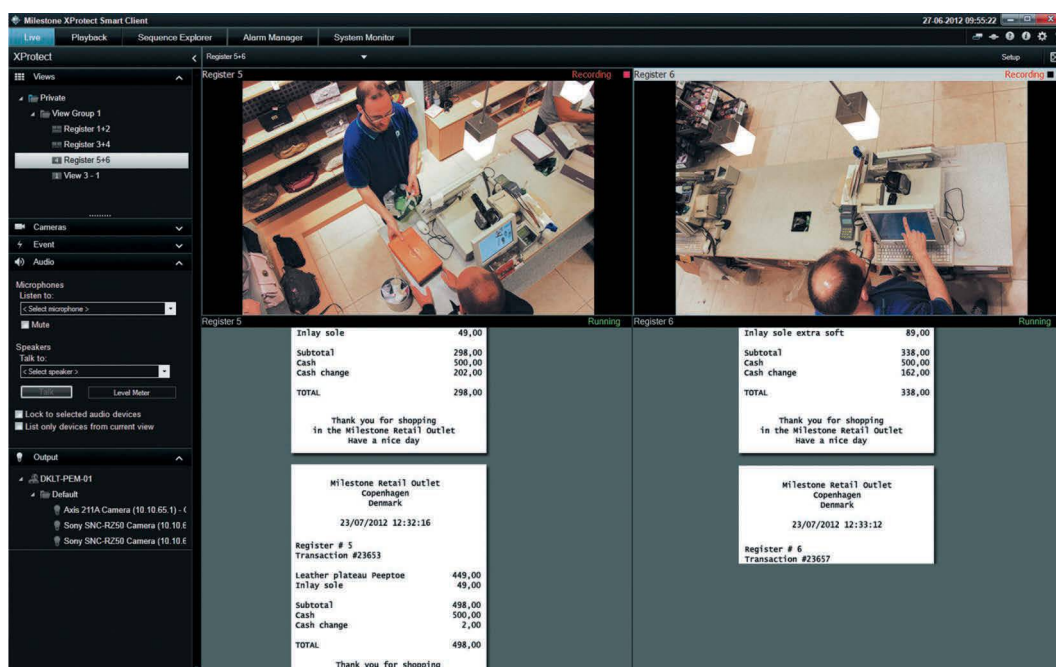
### 13.4.2 Point of sale

The introduction of network video in retail environments has made the integration of video with point-of-sale (POS) systems easier.

The integration enables the linkage of all cash register transactions to actual video footage of those transactions. It helps catch and prevent fraud and theft from employees and customers. It also makes it easier to search for and prove those suspicious activities. Through the captured video, managers can visually verify POS exceptions such as returns, manually entered values, line corrections, transaction cancellations, coworker purchases, discounts, specially tagged items, exchanges, and refunds. It can resolve questions such as whether the right amount was entered for the products placed on the counter, whether all items on the counter were scanned, whether a return was handled properly, whether an employee discount was given to a friend, and what a person who used a stolen credit card looked like. High-quality video from network cameras can provide the necessary information to identify and verify details such as the value of a bill or items handed to a cashier.

Some systems can catch POS exceptions and can store and show the receipts together with video clips of the events. In such cases, searching and viewing such events is possible as well (Figure 13.23).

A POS transaction or exception can be used to trigger a camera to record and tag the recording. For example, the opening of a cash register drawer can be used to trigger recordings. The scene prior to and following an event can be captured using pre- and postevent recording buffers. These event-driven recordings increase the quality of the recorded material, reduce storage requirements, and reduce the amount of time needed to search for incidents.



**Figure 13.23** An example of a point-of-sale system integrated with video surveillance. (Image courtesy of Milestone Systems, Brøndby, Denmark.)



A POS system with integrated video should help achieve the following:

- Good insight into the various payment transactions
- Reduction of internal and external shrinkage (theft)
- Preventive negative behavior among coworkers
- Awareness of how and when mistakes are made

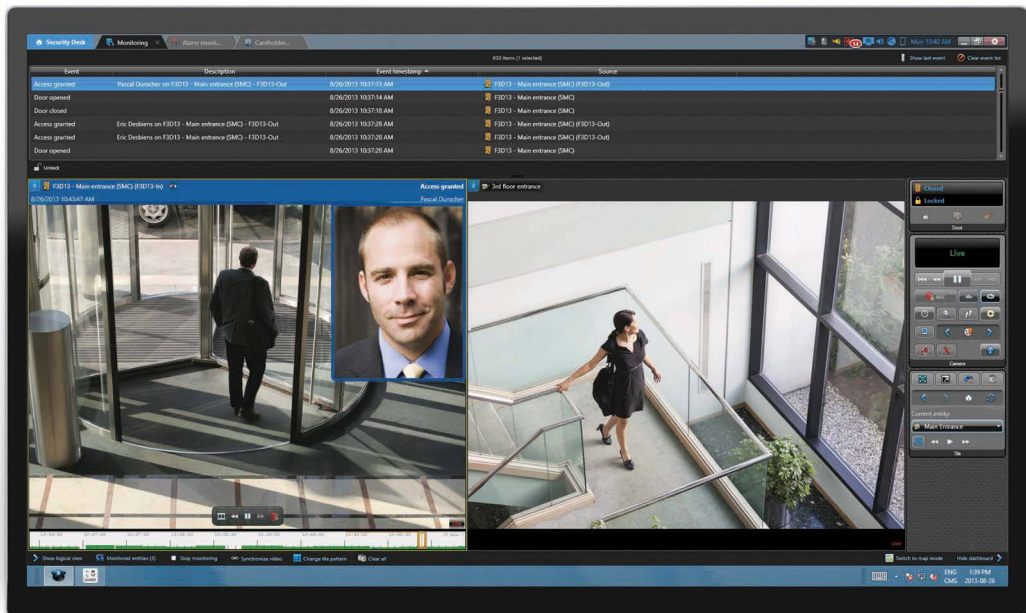
### 13.4.3 Physical access control

By integrating a video management system with a building's physical access control system (PACS), security managers can use video to log facility and room access (Figure 13.24). For example, when someone enters or exits a door, the door controller can trigger a camera to record the event. This way, suspicious activity can be verified and unwanted visitors identified.

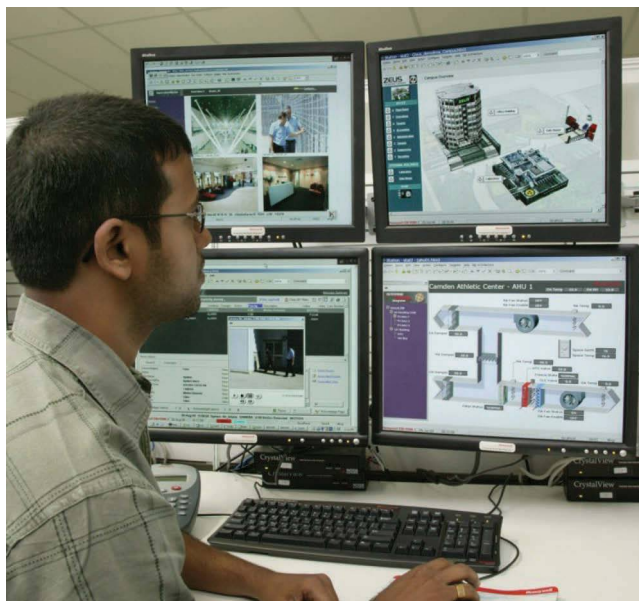
Video also makes it easier to identify tailgating, which occurs when the person who swipes their card knowingly or unknowingly allows others who did not swipe any card to enter via that door. If a problem occurs in an access-controlled area, it is important to be able to verify if any tailgating event occurred. To see if the system logged everyone who entered the area, the person investigating the incident would look at the recorded video and compare it with the PACS. For rapid identification of employees or visitors, the video surveillance operator can have access to all pictures in the badge system.

### 13.4.4 Building management

Video can integrate into a building management system (BMS) that controls a number of systems ranging from heating, ventilation, and air-conditioning to security, safety, energy, and fire alarm systems (Figure 13.25).



**Figure 13.24** An example of an integrated video surveillance and physical access control system. (Image courtesy of Genetec™, Montreal, Québec, Canada.)



**Figure 13.25** An example of a building management system that provides a single interface for monitoring a facility. The upper left-hand screen provides the operator with a quad view from different cameras. The lower left-hand screen provides an alarm summary with event-triggered video recordings. The top right-hand screen shows a campus graphic that makes it easy to navigate between locations. The lower right-hand screen shows an air handling unit and the status and values of different controls. (Image courtesy of Honeywell Building Solutions, Golden Valley, MN.)

Examples of integration include the following:

- In addition to activating alarms at the BMS, an equipment failure alarm can trigger a camera to show video to an operator.
- A fire alarm system can trigger a camera to monitor exit doors and start recordings.
- Analytics can be used to detect reverse flow of people into a building if a door is open or unsecured due to events such as evacuations.
- To save energy, VMD can be used to trigger lighting and heating systems to turn off when the room the camera is located is vacated.

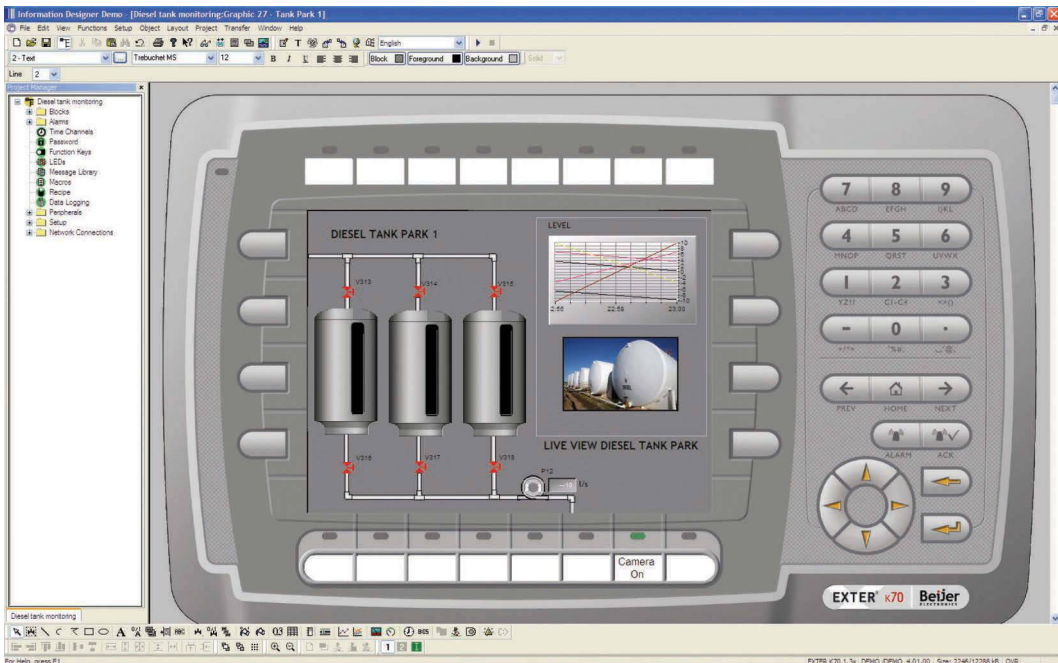
### 13.4.5 Industrial control systems

Remote visual verification is often required in complex industrial automation systems (Figure 13.26). By having access to the same interface for network video as for process monitoring, an operator does not have to leave the control panel to make visual process inspections. In addition, when an operation malfunctions, the control system can trigger the network camera to send images. In sensitive clean-room processes or in facilities with dangerous chemicals, video surveillance is the only safe or sustainable way to have visual access to a process. The same goes for electrical grid systems that have substations in very remote locations.

### 13.4.6 Radio-frequency identification

Tracking systems that involve radio-frequency identification (RFID) or similar methods are used in many applications to keep track of items. An example is luggage handling at airports that will keep track of luggage and direct it to the correct destination. When RFID is integrated with video surveillance, there is visual evidence when luggage is lost or damaged, and search routines can be optimized.





**Figure 13.26** Network video integrated with an industrial control system enables an operator to visually verify an activity remotely using the same user interface. (Image courtesy of Beijer Electronics, Malmö, Sweden.)

## 13.5 BEST PRACTICES

The video management software is the gateway to the entire video surveillance system. Today, several hundred different systems are available from different vendors. It is a challenge to choose the right one—the one that has just the right combination of features; is scalable; works on a platform that suits the business, the IT solution, and the staff situation; and is easy to use. There is a lot to consider, including the following issues:

- *Size of the system:* How many cameras and sites are there today, and what size do you foresee a few years ahead?
- *System architecture platform:* What platform best meets the needs? Is it an NVR- or a Windows-based hardware platform, an edge-based system, or a cloud solution? This choice is dependent on the collective requirements of the system.
- *One or many locations:* Is it a one-location or multiple-location business? How many devices are there in each location? How many operators will use the interface?
- *Storage and maintenance:* What are the recording and storage needs? Is there staff to maintain a hardware-based system? What is the cost of additional storage?
- *Scalability:* Some systems have limited scalability but are easy to install and operate, whereas others scale to thousands of cameras but may be complex to use in a small system. What is the current and estimated future device count? How many simultaneous camera views (split view) are needed? What are the licensing options? How many operators will use the system at the same time?
- *Complexity and functionality:* Is a basic system that enables recording and viewing of a few cameras sufficient? Or is an advanced system with, for example, event handling, mapping functionality, and support for integration with other systems required?
- *Integration:* Is integration with a POS system, a PACS, a BMS, or an industrial control system favorable? Look at possibilities of merging systems or support for adding other devices than video products. Does the software have an open API?

- *Analytics and intelligence:* Is the system required to have specific built-in analytics to help support the surveillance operations? In retail and transportation installations, segment-specific analytics such as people counting or license plate recognition might be desired.
- *Open or vendor specific:* Most network camera and video encoder vendors supply their own video management systems that are normally limited to only one brand of products. An open system from an independent company may provide better flexibility.
- *Activity level:* How active will the monitoring be? Will someone only check the video on a casual basis or only when an alarm goes off? Perhaps there will be dedicated staff to watch the video during all hours of the day. Different activity levels require different features, such as split screens, smart search, and effective alarm handling.
- *User level:* Who are the users, and what are their knowledge levels? Is it more important that the interface is simple and straightforward or that it is complex enough to meet every possible need? Ease of use is even more important if many operators are going to use it and their skills are diverse. Trained and skilled operators may have more detailed requirements on functionality and prefer a high-complex system. What are the operators' working patterns? Is a desktop client sufficient or do they need to be able to use mobile devices and web browsers as well?

# CHAPTER 14

## Hosted video solutions

As discussed in Chapter 13, there are many ways to manage video. The traditional way is to have a recorder, typically a digital video recorder (DVR) or network video recorder (NVR), in the same location as the cameras to manage the video. When IP cameras and video encoders are connected to an IP network that is connected to the internet, a new way of managing video opens up. Rather than storing it on-site, video can be managed and partly or wholly stored off-site in a data center. This way of video management is popularly referred to as hosted video. Hosted video, managed video services, cloud video surveillance, or Video Surveillance as a Service (VSaaS) refers to video surveillance provided as a web-based service. The National Institute of Standards and Technology (NIST) has defined the model of cloud computing. In summary, it is a service that is hosted and sold on demand, is accessed through the internet, and is elastic in the sense that at any given time its users, often referred to as subscribers, can use as much or little of the service as they want. Subscribers pay for and get a worry-free monthly service rather than procuring and maintaining a system.

In very broad terms, cloud services bring scalable video surveillance solutions to all-size organizations. Usually, only large organizations can afford to install the latest technologies and maintain on-site video surveillance systems, but for small businesses with multiple locations where each site has just a few cameras, hosted video solutions are ideal. Typical hosted video subscribers are convenience stores, gas stations, and small offices. By outsourcing the server, software, and maintenance to a service provider, organizations can limit their up-front investment. Instead of a large up-front capital expenditure (CAPEX), the user pays for the service on a monthly basis as an operating expenditure.

Though many use the term *hosted video* for any type of cloud video surveillance service, there are different types of surveillance as a service:

- *Managed video*: On-site recording and on-site video storage and remote management by the service provider
- *Hosted video*: Off-site recording, with data transfer to the service provider's site for storage and management

Often systems have both video streaming to the provider's site and storage at the camera site, for example, through onboard storage or network-attached storage (NAS). Nevertheless, these systems are still referred to as hosted video systems.

### 14.1 PRINCIPLES OF HOSTED VIDEO

Hosted video solutions brought modern network video technology to smaller installations. Suddenly, they could benefit from the HDTV resolutions, edge intelligence, and scalability of network camera technology that previously was only attainable by large organizations that had the

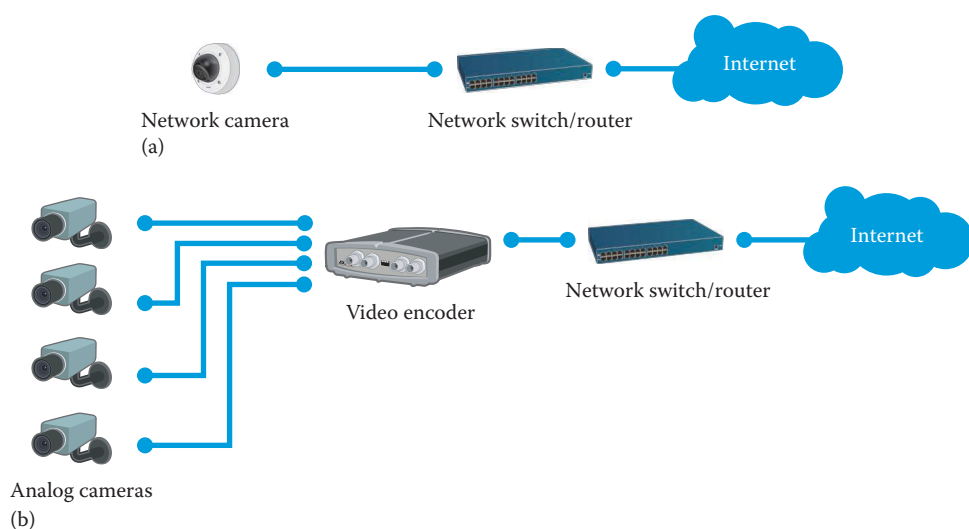
means to build advanced IT systems. Before, small systems would typically use an analog camera solution with a DVR or an NVR-and-IP-camera solution, which needed active management. With hosted video, these systems could grow one camera at a time while keeping investment costs to a minimum.

One of the keys to any electronic security deployment is ease of installation, as well as efficient ongoing system operation and maintenance. Hosted video solutions are available for a relatively small investment, so subscribers can limit their video surveillance infrastructure investments. They only need a camera or encoder and an internet connection (see Figure 14.1), and if they want to, they can let the system expand, add functionality, and subscribe to new services to meet the new challenges of a growing business.

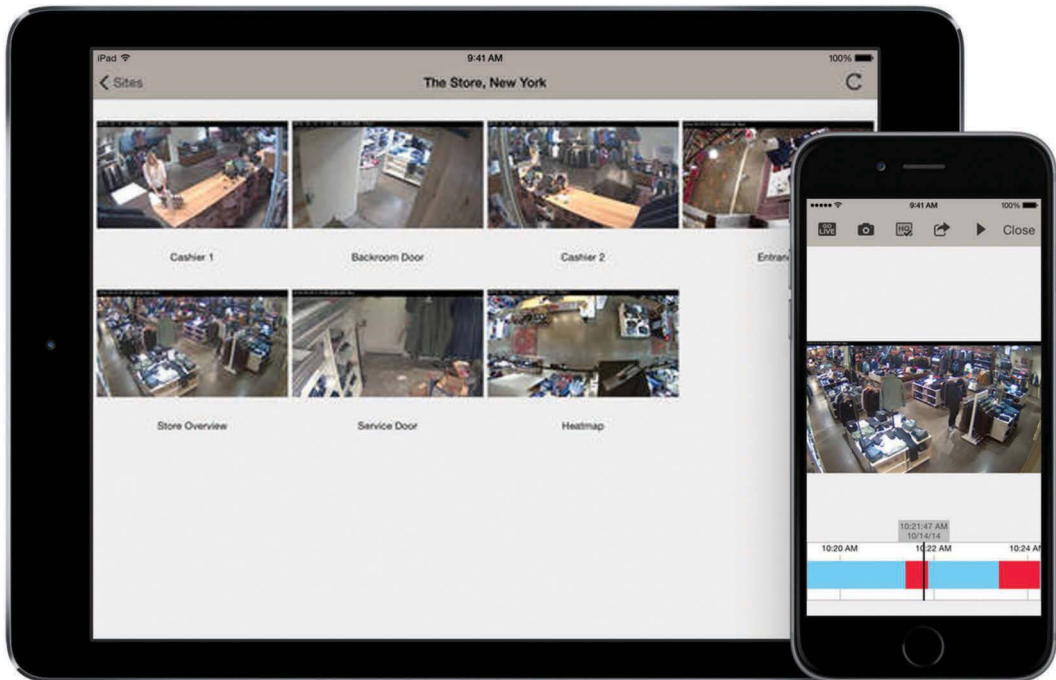
The subscriber's costs for the installation and maintenance of hosted video are lower than an NVR or PC server-based system for several reasons. There is no need to buy, install, or maintain a local recording and monitoring station because the video service provider (VSP) stores recorded data and maintains the system. Some providers even take care of the setup and network configuration. The system can grow one camera at a time as needs arise. If there are already analog cameras on a site, these can easily integrate into the new system. Video encoders transform the analog system into a hosted video solution that provides the benefits of network video while creating a future-proof and easily scalable solution. Another reason for low maintenance costs is that problems can be solved remotely in most cases. For example, there is no need to dispatch service technicians to adjust the focus when this can easily be done centrally and remotely.

Monitoring the video is easy and secure. Web clients and smartphone apps (see Figure 14.2) allow easy access to video from anywhere, and since viewing takes place over the internet, the subscriber does not have to invest in video management software. Live viewing, recordings, and notifications are available directly in the web browser.

The number and variety of cloud-based services is still growing, and much effort is going into establishing best practices for securing private, organizational, and governmental data. A benefit of storing video in the cloud is that data centers are experts at handling sensitive data. They continuously service the servers, keep them updated, and replace old hardware. Through video audit trails, records of access, and processes, they keep track of all the data and what happens to it. Many of the



**Figure 14.1** In its simplest form, the only equipment needed on a site is a network camera and an internet connection (a). Existing analog cameras can easily be integrated using a video encoder and an internet connection (b).



**Figure 14.2** One of the inherent benefits of hosted video is the ability to very easily monitor the video from any computer, tablet, or smartphone and go get alarms through simple text messages or push notifications. Here is an example of a cloud video viewing app for iPad® and iPhone.

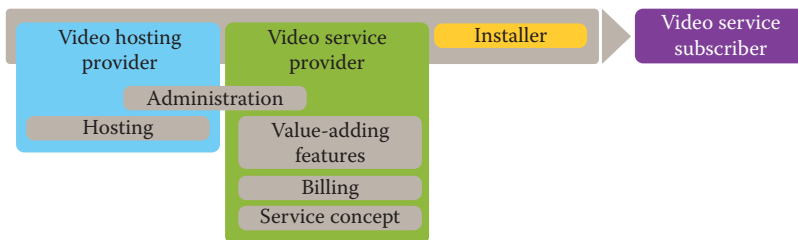
regulations and practices apply to hosted video as well. Therefore, subscribers can trust that their video data are safe. For more information about data security, see Section 14.5.

Hosted video can easily be combined with other systems to provide broader security and protection. For example, a hosted video solution can be integrated with a central station alarm platform. For more information, see Section 14.6.

## 14.2 STAKEHOLDERS OF HOSTED VIDEO

As shown in Figure 14.3, the main stakeholders of hosted video systems include:

- Video hosting provider (VHP)
- VSP
- Installer
- Video service subscriber



**Figure 14.3** Stakeholders of hosted video. The hosting provider is responsible for managing the video, while the service provider puts the service together, sometimes bundled with alarm monitoring, that is offered to the video service subscriber. The installer does the physical installation on-site and is often contracted by the video service provider.

The VHP, the VSP, and the installer play different roles in providing the complete solution and the appropriate services to the video service subscriber. Sometimes, one company can take on several of those roles. Many alarm operators also offer hosted video services. Sometimes, the user sets up their own cloud solution to support many different facilities. This is called a private cloud solution, as opposed to the public cloud solution that is usually implied when talking about hosted video surveillance.

### 14.2.1 Video hosting provider

The VHP is responsible for running the actual system that connects and manages all the cameras and the video stored in the hosted video system. Some hosting providers run their own data center, but many rent space in a big data center run by companies such as Amazon, Google, and Microsoft®.

The VHP supplies the hosting service to the VSP and guarantees data backup, recovery, performance, system availability, and security. Providers must demonstrate that their servers have the necessary security protocols to protect private and confidential data. Typically, the hosting provider runs a scalable and cost-efficient operation and sells services to many service providers.

### 14.2.2 Video service provider

The VSP buys the video hosting from the VHP, provides a new interface, and sometimes integrates the video with other systems. For examples of integration with other systems, see Section 14.6.

The service provider is often the stakeholder that has the commercial relationship with the video service subscriber and charges a monthly service fee for the video surveillance service. The monthly fee varies depending on the number of cameras per site and other factors such as frame rate, resolution, and retention, all of which affect the amount of storage and bandwidth needed.

The service provider also charges for the value provided to the subscriber, such as active, scheduled, and event-triggered monitoring of camera feeds. Trained operators can use special monitoring software to monitor video and take action on anything that happens. Operators can send alarms, save recordings and images, and notify system users of events. The more services the user wants and the more of an active role they need operators to take, the larger the fee. Service providers refer to this business model as recurring monthly revenue.

In other words, a subscriber can get a number of different services from a service provider. If the service provider is a security company, they get security and safety as well as a video management solution. Storeowners who previously had to manage their own surveillance system can pay a fee to have someone manage it for them. Through their computer, tablet, or smartphone, they can access the system remotely and, for example, do a forensic search for a stolen item.

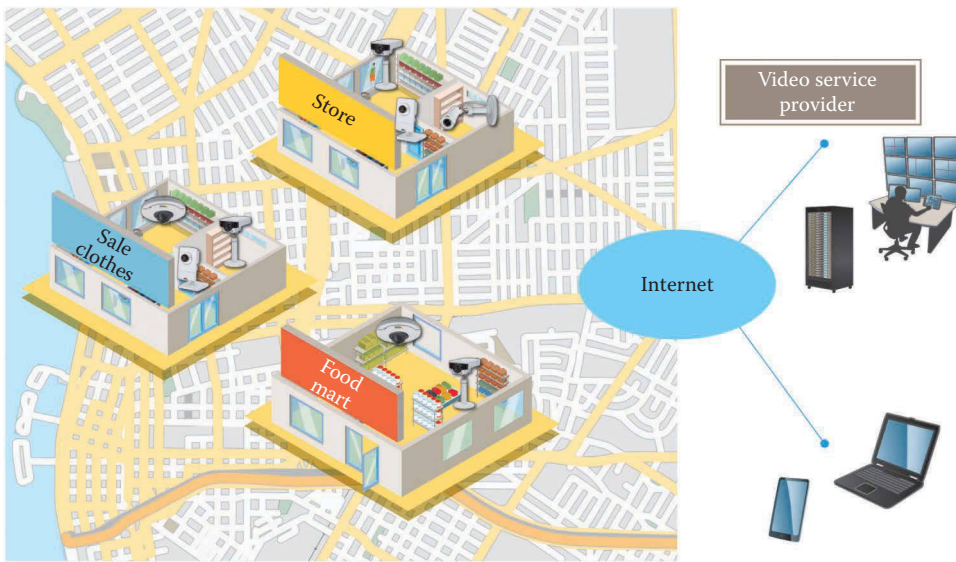
### 14.2.3 Installer

The installer does the actual installation on the different sites. For installers, hosted video is easier to install than other types of surveillance systems, partly because they need very little equipment on-site. The installation is even easier if the cameras are staged (i.e., preconfigured and connected to the service provider's system) in a staging facility before the installer gets them. On-site, the installer only has to install the cameras, connect them to the network, and point them in the right direction. Many cameras have remote focus, which means that the user can adjust focus and field of view without leaving the office.

### 14.2.4 Video service subscriber

Hosting video in the cloud provides a simple, affordable, and scalable recording solution for video surveillance. This is a viable alternative, whether it is a small or midsize business with one store to monitor remotely, a franchise with multiple sites that need simultaneous monitoring, or a large corporation looking for secure off-site archiving to satisfy internal policies or compliance issues.





**Figure 14.4** A typical hosted video setup in a small retail setting.

As with other cloud services, a hosted video surveillance platform allows users to leverage their system's scalability and cost effectiveness. Because the need for on-site storage and maintenance is greatly reduced or eliminated, business owners can reduce both the capital investment in physical security technology and the total cost of ownership. There is great value in that video service subscribers can rely on their service providers to help them grow elastically. The service providers always make sure that they have the resources that serve the purpose best; they replace hardware and update software. From the subscribers' point of view, they are buying a future-proof and worry-free solution. Figure 14.4 shows one of the most common hosted video setups: a retailer who subscribes to hosted video and managed security services and who uses a laptop and a mobile device to view video and manage their system.

The storage, computing, infrastructure, and security aspects of the hosted environment transfer many responsibilities to the service provider, including:

- *Personnel* to manage the solution and take care of demands, scheduled maintenance, and unforeseen events
- *Financing and depreciation* of storage, computing, infrastructure, and security applications
- *Recurring facility costs* for providing hosting of video, for example, electricity, HVAC, environmental monitoring, and loss prevention
- *Continuous and periodic assessment* to determine storage, computing, cyber security, and space limitations as demand rises and falls

The flexibility of hosted video makes it ideal for surveillance that is temporary by nature. Public events such as carnivals, marathons, demonstrations, and outdoor concerts have very different needs and altogether different types of infrastructure than a parking garage or government building. Video surveillance systems for public events must be deployed quickly, be easily accessible, and provide the ability to manage video in a scalable, secure, and easy way for a short time. Just as quickly, they need to be removed from the site again. Some organizations need surveillance that can be moved frequently to meet changing site needs.

There are also enterprises that currently operate large on-site surveillance systems but that, because they cannot afford to lose any critical video data, have identified a need to move off-site to a secure cloud-based location. Cash rooms, server rooms, and pharmacies are examples of areas that have

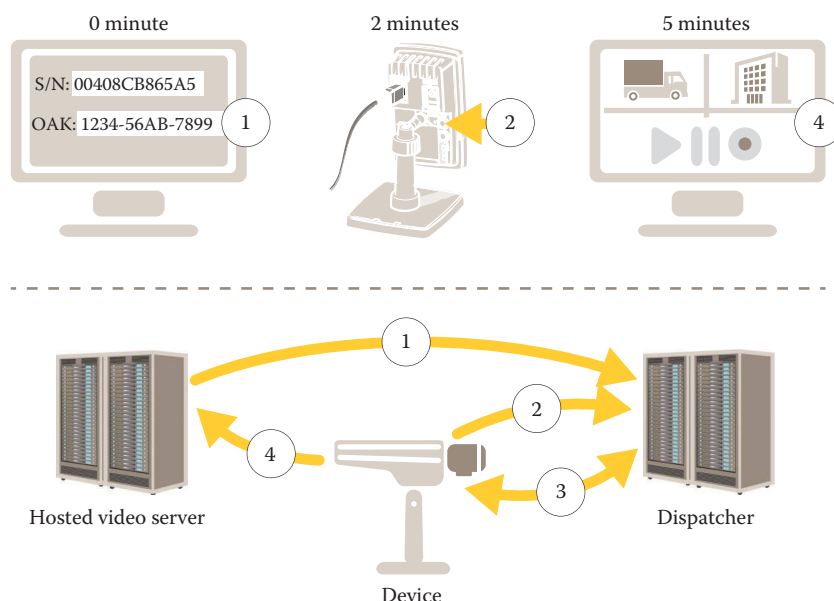
critical video content that could be very problematic for the business to lose. Enterprises can isolate these cameras and simply host them off-site, so the video is secure and no longer at risk of loss due to video manipulation, local recorder failure, or theft.

### 14.3 SETTING UP HOSTED VIDEO

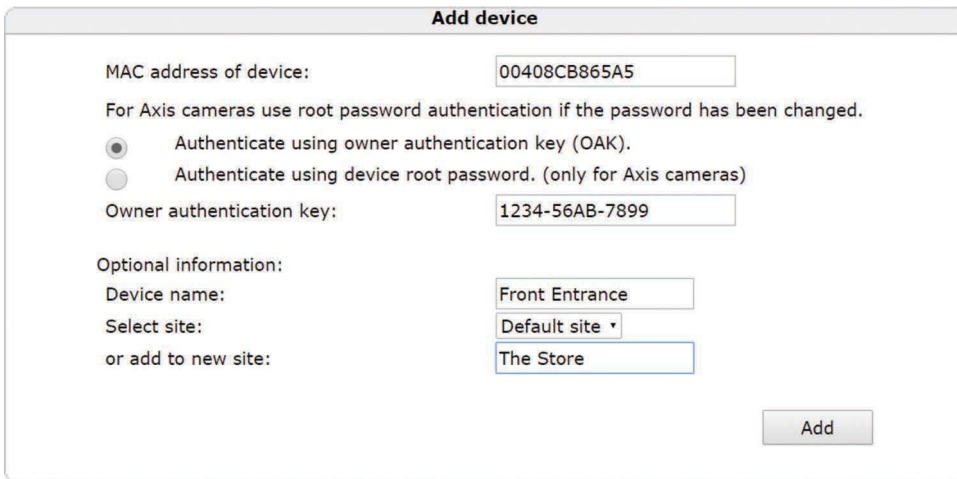
After installation at the site, the network camera or video encoder has to connect to the hosted video service. Getting the video device to communicate through the router and firewall at the remote site sometimes requires advanced networking knowledge. Although it can be a complex and daunting process, it does not have to be. Some hosted video systems focus on ease of install. For example, a camera can have a push button that simplifies connection to a hosted video service. When the button is pushed, the camera initiates and establishes the connection. Then the camera can be activated by entering the MAC address, which is the same as the serial number, and the owner authentication key into the system. The key is supplied with the product and consists of a series of numbers and letters that only works with the product with which it is associated. Figures 14.5 and 14.6 explain how the connection procedure works.

Rules for how and when the system should record video are based on user requirements. The video feeds are set up to record at specific resolutions and frames per second. They can also be configured to record at specific times or under specific conditions (see Figure 14.7). For example, external contacts and sensors, built-in passive infrared (PIR) sensors, and video motion detection can be used to trigger recordings.

One of the challenges with hosted video is bandwidth, especially in scenarios that need high frame rates, high resolutions, or several cameras per site. Local storage (sometimes referred to as edge storage), such as a NAS or an SD card, can be used to overcome bandwidth limitations. Edge storage means that the video is recorded and stored locally – near or onboard the camera. NAS devices are placed on-site and connected to the network, whereas SD cards are usually inserted directly



**Figure 14.5** Some cameras are primed for easy configuration of hosted video. Here is an example of a camera with a push-button solution that uses a three-step configuration: (1) Open the web interface of the cloud software and enter the serial number and owner authentication key. (2) Press the camera's button. (3) The camera talks to the dispatch server and the hosting provider's server to establish the connection. (4) The camera can start recording to the hosting provider's server.



**Add device**

MAC address of device:

For Axis cameras use root password authentication if the password has been changed.

☒ Authenticate using owner authentication key (OAK).  
☐ Authenticate using device root password. (only for Axis cameras)

Owner authentication key:

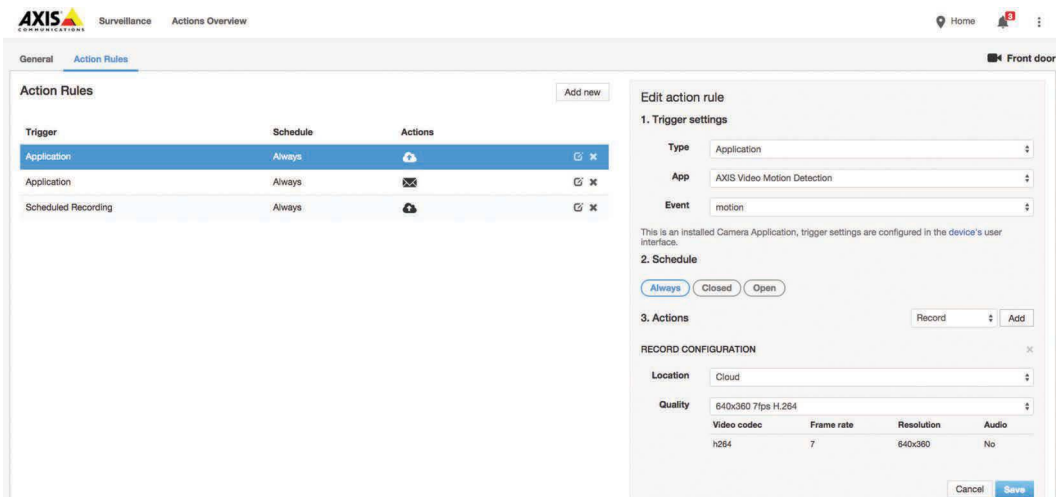
Optional information:

Device name:

Select site:

or add to new site:

**Figure 14.6** Adding a device to a hosted video system.



**AXIS** Surveillance Actions Overview

Home

**General** **Action Rules**

**Action Rules**

Trigger	Schedule	Actions
Application	Always	
Application	Always	
Scheduled Recording	Always	

**Edit action rule**

1. Trigger settings

Type:

App:

Event:

This is an installed Camera Application, trigger settings are configured in the device's user interface.

2. Schedule

☒ Always ☐ Closed ☐ Open

3. Actions

**RECORD CONFIGURATION**

Location:

Quality:

Video codec	Frame rate	Resolution	Audio
h264	7	640x360	No

**Figure 14.7** Setting up scheduled and triggered recordings.

into the camera or video encoder. Local storage also provides redundancy if the network connection goes down or becomes overloaded by other data transmissions. A scalable network video system has the ability to send several video streams from each camera, which makes it possible to simultaneously record video to local and remote storage, as well as view live video on mobile devices. Each stream has a different resolution and frame rate, depending on storage type, bandwidth, and viewing modes (see Figure 14.8).

## 14.4 CHARACTERISTICS OF HOSTED VIDEO

A hosted video management system (VMS) has many of the same functions as a traditional server-based VMS:

- *Observation:* Continuous and event-based monitoring of specific video sources
- *Forensic review and verification:* Searching for video associated with intrusion, safety, or environmental alarms

### About stream profiles

Stream profiles are used when fetching video streams from devices. The stream profile defines properties of the video stream, such as resolution, framerate, codec etc. When configuring alarm recordings, scheduled recordings, live view etc., a stream profile needs to be specified along with the configuration. A single stream profile can be used by multiple recording configurations at the same time, as well as live streams.

[Read more](#)

### Manage stream profiles

Name / Optimized for	Device	Resolution	Framerate	Videocodec	Overlay	Audio	Delete
Local storage	00408CB865A5	640x480	30	h264	%G-%m-%d %T	No	<a href="#">Delete</a>
Cloud storage	00408CB865A5	640x480	7	h264	%G-%m-%d %T	No	<a href="#">Delete</a>
Live viewing non-IE	00408CB865A5	640x480	7	jpeg	%G-%m-%d %T	No	<a href="#">Delete</a>
Live viewing IE	00408CB865A5	640x480	7	h264	%G-%m-%d %T	No	<a href="#">Delete</a>
Web	00408CB865A5	640x480	7	h264	%G-%m-%d %T	No	<a href="#">Delete</a>

### Create stream profile

Name / Optimized for:

Resolution: 640x480

Framerate: 30

Videocodec: h264

Overlay: Date and time

Create

Internet Explorer

Other browser

Live viewing

Manual recordings

Scheduled recordings

Alarm recordings

Cloud storage

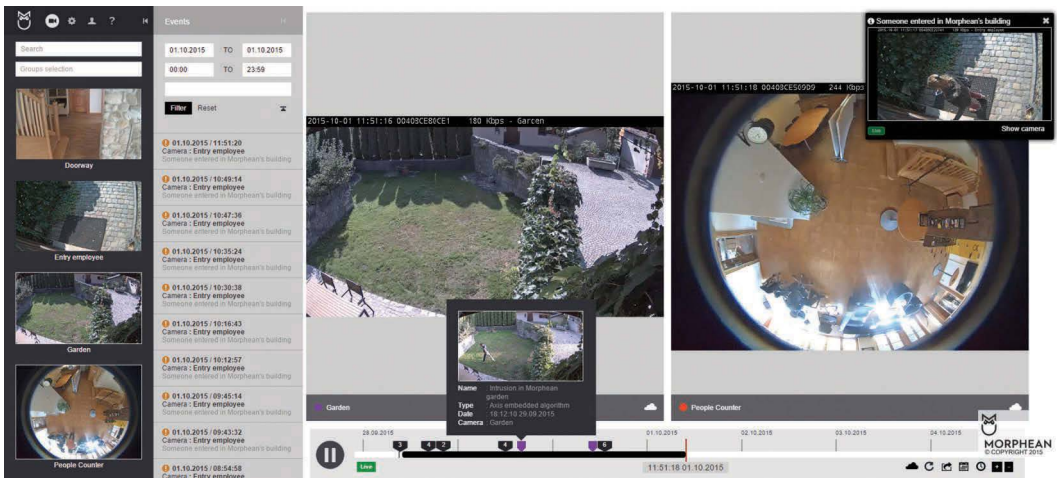
Local storage

**Figure 14.8** Setting up stream profiles with different frame rates and resolutions for different types of storage and viewing modes.

- *Multistreaming*: Automatic video quality selection dependent on user location to provide an optimized user experience, for example, lower-resolution video for remote viewing and high-resolution video when accessing the system on-site
- *Export of video clips*: Download of video material for use as evidence in a court of law
- *Detection and identification*: Monitoring of specific video sources like cameras equipped with specialized video analytics such as perimeter, crossline, or motion detection
- *System integration*: Integration with alarm monitoring software and hosted access control
- *Business intelligence*: Integration of business intelligence-focused video analytics such as people counters, queue monitoring, and heat mapping to assist the user in optimizing their business
- *System maintenance*: Monitoring the health, status, firmware version, storage, and available resources of the connected network cameras and encoders

So why choose hosted video? It offers similar functionality to a traditional surveillance system with little up-front investment. Hosted video brings flexible and low-maintenance storage options. Hosted video subscribers can combine cloud and local storage with less trouble and less cost than in traditional surveillance systems. In most designs, the costs and maintenance associated with legacy DVRs, NVRs, or VMS server are no longer necessary. As part of the service, video is recorded in the cloud, so the system is not dependent on local recorders that need to be maintained and that can be stolen or confiscated. On the flip side, if a specific site has more than 10 cameras, a local recorder is typically more cost efficient. With more than 20 cameras, a client-server system has better features and provides better scalability.

The interface is one of the areas in which hosted video management software still stands out against many video management software applications. It is usually web-based or specially developed for mobile devices, and it focuses on basic functionality and ease of use (see Figure 14.9). Users need only open a web browser or a smartphone or tablet app and remember their username and password. For the user, maintenance is also problem-free as it all happens behind the scenes. The hosting provider and the service provider take care of the system and handle all upgrades.



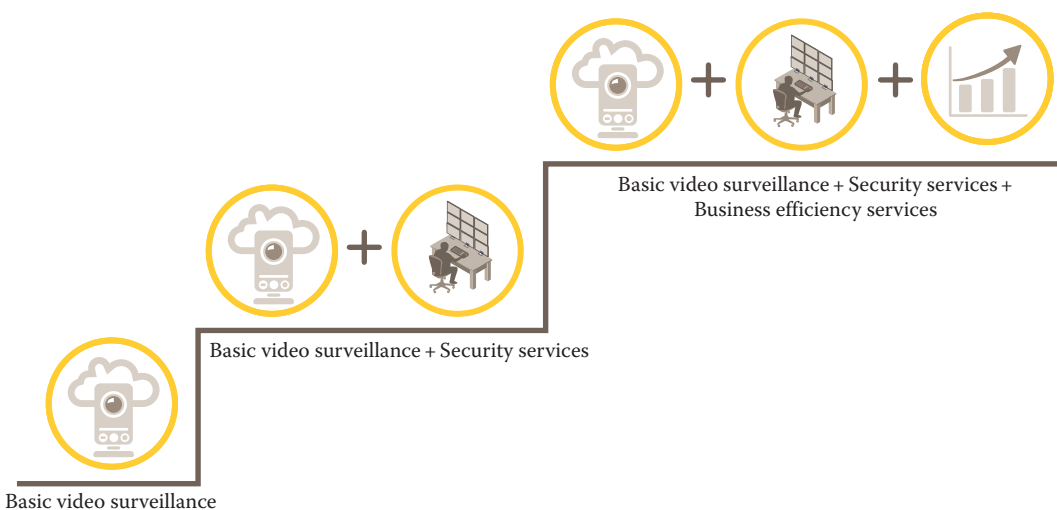
**Figure 14.9** One of the characteristics of most hosted video applications is a clean, functional, and modern-looking interface that is easy to use. The focus is on system overviews, such as split screens and thumbnail sliders, timelines, event searches and alarm management, pop-ups, and video export. (Image courtesy of Morphean, Granges-Paccot, Switzerland.)

### 14.4.1 Different needs, different services

The first step in finding the right video surveillance solution should be to identify what the user wants and what the purpose of the system is. Hosted video can meet many different levels of video surveillance needs. It is up to the VSP (and in some cases the security company) to package solutions that meet those requirements. Hosted video subscribers want peace of mind. A good solution provider should also offer a great deal of flexibility so that subscribers can add services as their challenges change and their business grows. Figure 14.10 illustrates a ladder of video surveillance needs. At each level, different preventive, forensic, or comfort levels can be met.

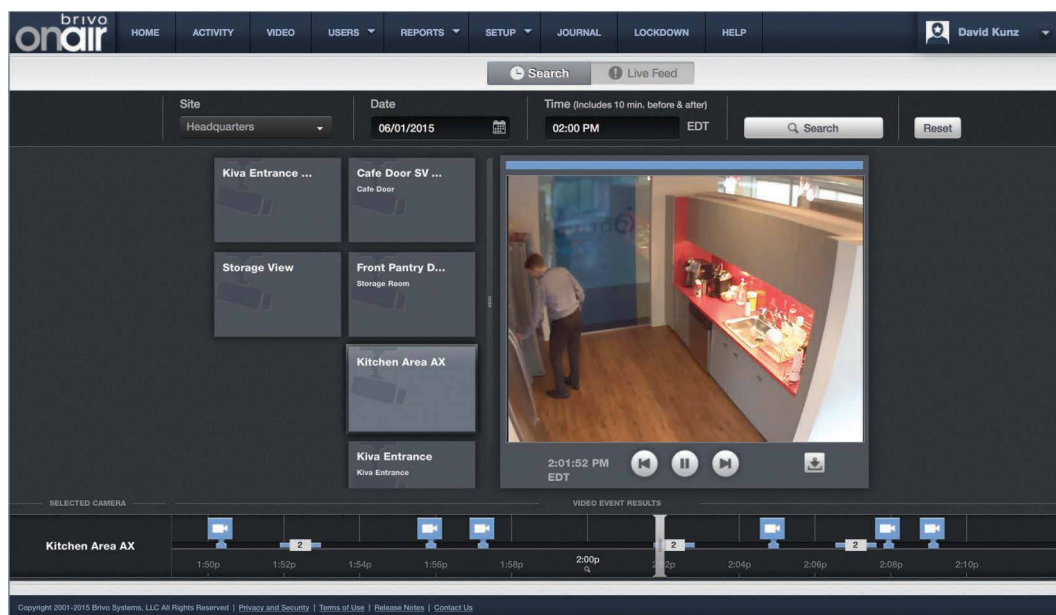
#### 14.4.1.1 Basic video surveillance

Subscribers' main objective might be to protect property and prevent loss. They probably also want to feel safe and confident that the system has them covered. They find comfort in that the cameras



**Figure 14.10** A ladder of video surveillance needs, where each step adds more services. It begins with basic video surveillance, advances to security services (such as alarm monitoring), and finally services that help make the subscriber's business more efficient.





**Figure 14.11** The user interface of a hosted video application. The timeline (bottom) and the search function (top) let the user browse between triggered and scheduled recordings, select recordings to export to a flash drive or other storage device, and switch to live view. To select or switch between cameras, the user only has to click on the thumbnails or labels (main area/middle). (Image courtesy of Brivo, Bethesda, MD.)

make staff, managers, and business owners feel safe and that they can get an overview of the premises and share video easily from any location (see Figure 14.11). They know that they easily can follow up on incidents and perform basic forensic investigations. Storing video is easy; it is available in the cloud and can be stored locally by plugging in a NAS or inserting an SD card. If subscribers need it or government agencies demand it, the cloud provides theft-proof storage of critical data.

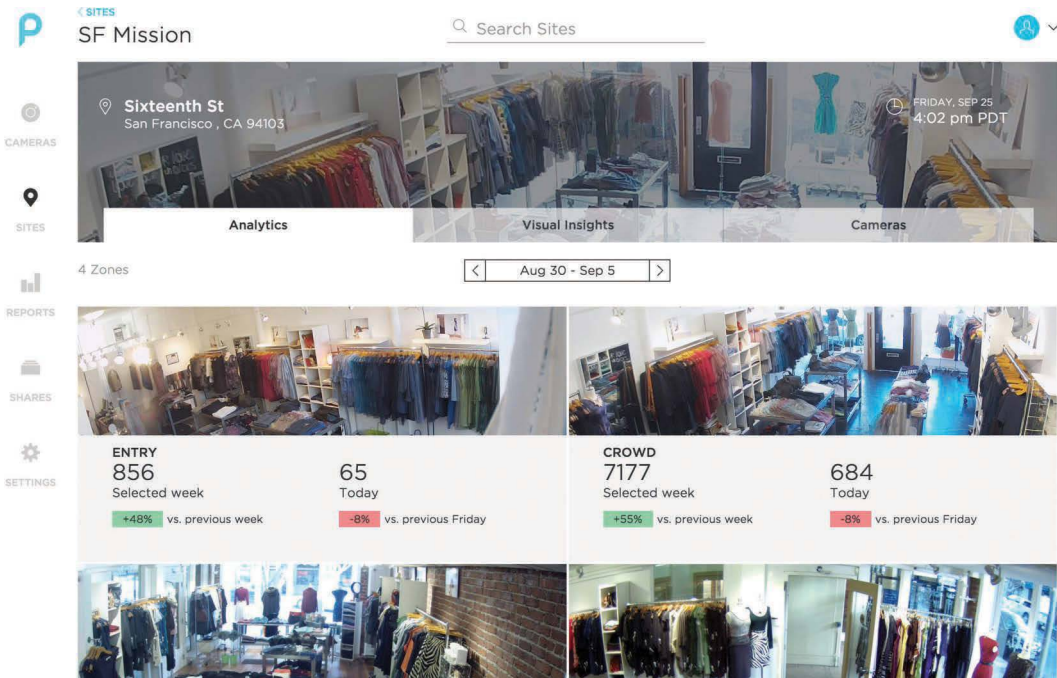
#### 14.4.1.2 Alarm monitoring

Some subscribers find comfort in knowing that there are trained operators who can verify activity through video, making visitor and contractor entries more efficient and secure. When an alarm occurs, security personnel will act on it. Users can even watch the whole scenario remotely if they want to. Security operators and automated guard forces work together to monitor the system and keep an eye on the premises, daily deliveries, cash registers, and deposits. The system helps hosted video subscribers and security centers make better decisions on manpower distribution, and it allows them to centralize the guard force even if they need to guard several locations and buildings. A shopping mall operator might want to deploy a video system and provide it as a service to the tenants.

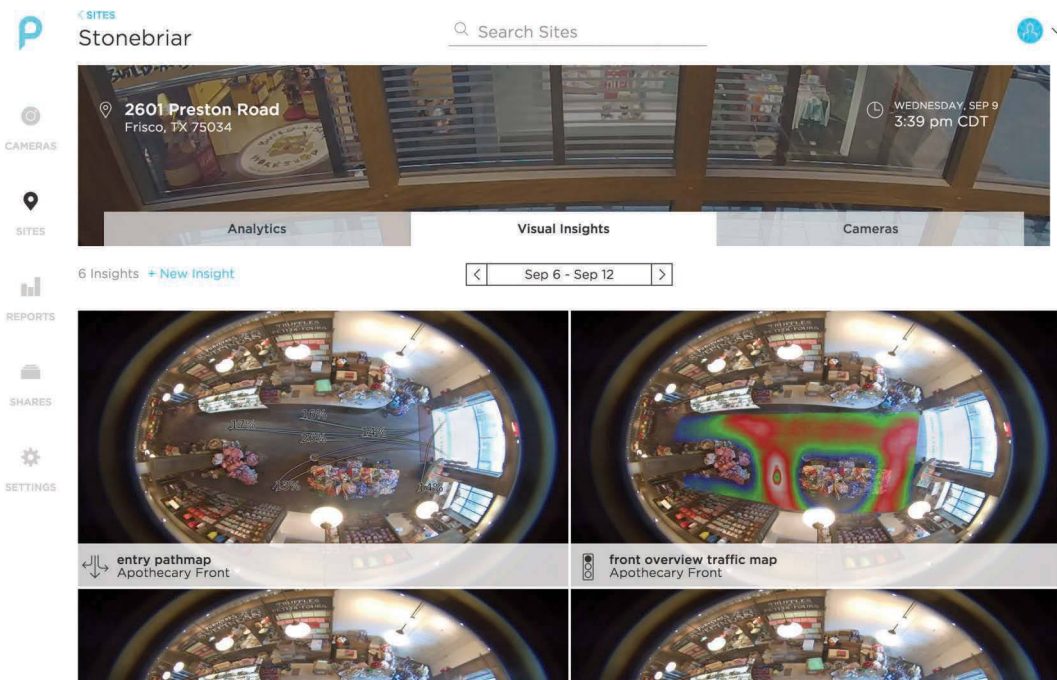
#### 14.4.1.3 Business intelligence

Many business owners have understood that because its systems are easy to scale and deploy, even temporarily, hosted video can help them build a more successful business. Critical data are safely stored in the cloud and are easily shared with the police and judicial system. Through video, they can monitor campaigns, find ways to make merchandizing stand out more, make sure that the opening and closing hours are kept to, as well as look at the status of different areas. They can use analytic tools that help them resolve incidents faster; predict behaviors to prevent inventory loss; follow movements through heat mapping, customer traffic, and dwell time monitoring; identify patterns based on parameters such as gender; and use people counting as scheduling input and as a tool for evaluating merchandise positioning (see Figures 14.12 and 14.13). To read more about analytics for business intelligence, see Chapter 16.





**Figure 14.12** An example of a cloud-based solution for people counting and customer traffic (crowd) monitoring. (Image courtesy of Prism Skylabs, San Francisco, CA.)



**Figure 14.13** An example of a cloud-based solution for customer traffic (crowd) monitoring and heat mapping. The path map (left) shows how the percentage of customers walking in each direction and how far they go. The heat map (right) shows where the hot zones in the store area are, that is, which areas have the most traffic. (Image courtesy of Prism Skylabs, San Francisco, CA.)

If they integrate video analytics with the point-of-sale system (POS), storeowners can multiply the benefits, for example, by having cash transactions trigger recordings and tag them with images of the receipts.

## 14.5 DATA SECURITY

Data security is top of mind for businesses, government, and individuals alike. Possible risks include compromised video integrity through manipulation of the video images, the breach of information systems to which the hosted video system is connected, and denial of service since the network camera will only support a finite number of users directly. There is also the risk of someone taking control of the device itself, to destroy onboard data or to redirect the video stream elsewhere, which in turn can lead to people gaining unauthorized access to protected areas.

Businesses and governments around the world depend on the internet and cloud services. Over the years, many technologies have become available for securing IT environments. Hosted video is repurposing many of those technologies. Any decent hosted video solution should meet the following security standards:

- *Two-factor authentication*: The user must provide a combination of things only the user knows, such as a username, a strong password, and an RSA SecureID token (e.g., a USB dongle, smart card, or key fob). This prevents wrongful use of leaked usernames and passwords.
- *Encryption of passwords*: Usernames and passwords are always transferred encrypted between the user and the hosted system. In addition, the hosting provider stores the password in an encrypted format in combination with a technique known as salted password hashing, which ensures that nobody can use the information even if the password database is compromised.
- *Encryption of all streams*: The video and data connections are always encrypted to prevent other parties from getting access to the material.
- *Machine-to-machine authentication*: The connections between cameras and hosted video servers are authenticated using signed certificate technology to prevent man-in-the-middle attacks.
- *Camera-owner authentication*: To avoid wrongful use of deployed cameras, it is only possible to connect a camera to the hosted service using a unique identifier that only the camera owner knows.

In many ways, storing video data in the cloud can be more secure than storing it exclusively on a DVR, NVR, or local server. With hosted video, video is not just stored in on-site recording devices, but some or all can also be stored and managed in the cloud. This way, video evidence is always protected against theft and vandalism. With the system redundancy provided by inexpensive redundant storage devices, such as NAS or SD cards, recordings are protected even if the network goes down.

### 14.5.1 Audits, laws, and certifications

The same compliance regulations met by the big data storage players also apply to VHPs. Data centers should be audited, and it is also possible to attain certification that guarantees compliance with a best-practice framework for secure information management. The following section outlines some of the relevant standards.

#### 14.5.1.1 Standards for Attestation Engagements

Statement on Standards for Attestation Engagements No. 16 (SSAE 16) is put forward by the Auditing Standards Board of the American Institute of Certified Public Accountants. SSAE 16 replaced Statement on Auditing Standards No. 70 (SAS 70) in 2011. The reasons behind the new standard include the intention to align with the global standard for reporting on controls at service organizations, International Standard on Assurance Engagements (ISAE) No. 3402, and to overcome the limitations of SAS 70. SSAE 16 regulates how service companies report on compliance

controls and, unlike SAS 70, includes a detailed description of the system, its services, processes, policies, procedures, staff, and operational activities, as well as a written statement of assertion. ISAE No. 3402, Assurance Reports on Controls at a Service Organization, was issued in 2009 by the International Auditing and Assurance Standards Board, which is part of the International Federation of Accountants. Although SSAE 16 and ISAE No. 3402 align well with each other, service organizations with global operations have good reason to examine both standards.

#### 14.5.1.2 ISO/IEC 27001 standard

ISO/IEC 27001 is part of the international ISO/IEC 27000 standard series on information technology. It was revised in 2013 and helps data centers and other organizations to identify and treat security risks and to set up tools for managing information security management systems. The requirements are generic and intended for all organizations, regardless of type or size. ISO/IEC 27001 is a best-practice framework and certification is not mandatory. However, gaining an official certification by an accredited certification body can give service organizations paybacks such as reductions in incidents, faster system recovery, greater security awareness, more credibility, and status as preferred supplier.

#### 14.5.1.3 Federal Information Security Management Act

The Federal Information Security Management Act (FISMA) is a U.S. federal law passed in 2002. Federal agencies and their contractors must follow the act, which regulates information system policies, procedures, and practices. In essence, it is a collection of standardized best practices for strengthening information security and reducing security risks. Companies dealing with federal agencies can choose to outsource to a FISMA-compliant data center or try for on-site compliance. On assignment by FISMA, the NIST is responsible for developing standards and guidelines for choosing, categorizing, and evaluating non-national security federal information systems. FISMA also requires annual reviews of the effectiveness of the agency's information security program and that the result is reported to the Office of Management and Budget (OMB). The OMB uses this information to prepare an annual congressional report on agency compliance with FISMA.

#### 14.5.1.4 European Union Agency for Network and Information Security

The European Union Agency for Network and Information Security (ENISA) is a body of expertise that helps the European Commission (EC), its member states, and the business community to address, respond to, and prevent network and information security problems. Part of this work is to assist the EC with legislative preparations in the field of network and information security. ENISA does not inspect, regulate, or enforce information security laws. Both legislation and enforcement are handled on a European and national level through EU bodies, such as the European Data Protection Supervisor and Europol, and national agencies.

## 14.6 INTEGRATION WITH OTHER SYSTEMS

---

While hosted video is most commonly used as a stand-alone application, it can become even more powerful when integrated with other systems. Hosted solutions are very dynamic and make new functionality and services easy to approach. For example, the subscriber starts out by signing up for video services and then adds access control. In the next step, they might add video analytics, such as heat mapping or people counting to certain cameras, or perhaps something more innovative, such as biometric analytics. Many services can be added remotely without even having to go to the site.

The most common systems to integrate with are central station automation platforms used in alarm monitoring and hosted physical access control systems.

### 14.6.1 Integration with central station automation

The alarm monitoring industry has been around for many years. It is based on monitoring remote alarms, which are often simple contacts or PIR sensors that detect motion. When an event sets



**Figure 14.14** Integrated central station software and a hosted video solution can provide a very effective solution for the operator to decide if the alarm is an event that calls for dispatching the police or not.

off an alarm, a traditional central station has no other information than the alarm to act on. For a long time, video verification of the alarm was discussed without producing a good scalable solution.

For many of these scenarios, hosted video can be a great solution. To make the solution efficient, the station automation software and the hosted video system should be integrated so that the operator can use the same interface to get information and file reports.

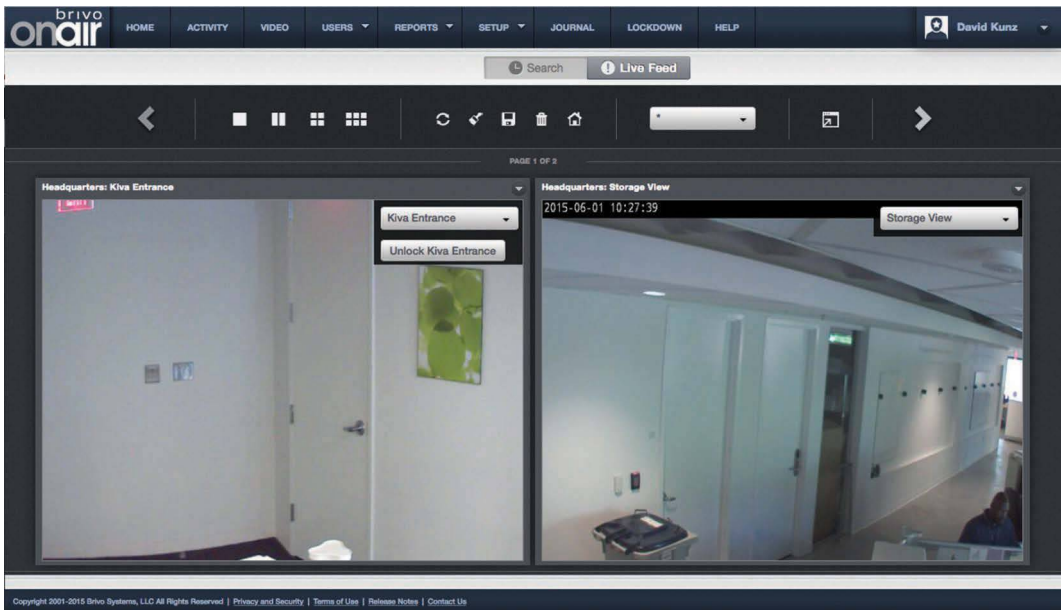
As the use of standard phone lines (POTS) is in decline in favor of IP-based communications, central stations look for systems, such as cloud video surveillance solutions, that can use the same network (Figure 14.14).

### 14.6.2 Integration with physical access control

Hosted access control has been around for a number of years and has become an integral part of the security industry. One of the reasons it was adopted so fast compared to video surveillance is that it involves a lot less data traffic, so the dependency on bandwidth is much lower. The benefits of an outsourced model for access control are the same as with video, and time has shown that the data security measures are sufficient also for physical security applications.

Video can add another dimension to the access control system and is a new business opportunity for the provider of the hosted access control system. The benefit for the users is that they can get visual verification for all access control transactions, along with general video surveillance (see Figure 14.15).





**Figure 14.15** By integrating hosted access control as well with video, the user will have one system covering several physical security needs. Users can search recordings, save recordings, and switch between cameras and split screens (top). While watching the live video stream (Live Feed), users can identify visitors and click a button in the interface (Unlock Kiva Entrance) to unlock the door. (Image courtesy of Brivo, Bethesda, MD.)

## 14.7 BEST PRACTICES

Hosted video is a cloud service with some unique characteristics that calls for careful considerations before deciding if it is the right solution:

- *Scattered locations:* Typically, hosted video solutions have the most benefits when there are many different locations and only a few cameras are installed per location.
- *Full cloud services:* Some vendors refer to remote access to a DVR or NVR as a cloud-based service. While the internet is certainly used for accessing them, the recordings are not actually stored in the cloud. With this solution, you miss out on many of the benefits of a true hosted video solution.
- *Off-site redundancy:* The need for a redundant recording solution with video recorded off-site is often a driving factor for a hosted solution.
- *On-site redundancy:* Different local storage solutions can be used to limit the need for bandwidth and create redundancy, prevent data loss in the case of a network disruption or unreliable bandwidth, and add recognition or identification capabilities to a hosted video verification system.
- *Multisite redundancy:* Large amounts of data (i.e., high-quality video or long recordings) affect storage and archiving requirements. Some users need a redundant system with both remote and on-site storage.
- *Bandwidth estimates:* Bandwidth is an important factor when creating a hosted video solution. For the results to be useful, first determine what the requirements on image quality are. Detection needs less bandwidth than recognition or identification. High-resolution images need more upstream bandwidth than low-resolution images.
- *Bandwidth handling:* Some system providers offer hardware devices and software that help manage streams, recordings, and storage while consuming less bandwidth.
- *Security:* Any serious hosting provider should be able to prove that their data storage solutions are secure. They should use approved authentication processes and comply with the latest

security standards. Service providers also need to employ security measures, partly by making sure the hosting providers follow best practices so that data are always streamed, recorded, and stored without risk of eavesdropping.

- *Alarm monitoring:* Consider alarm monitoring services for increased comfort and help with verifying and acting on alarms, centralizing guard forces, and making the right decision at the right time.
- *Integration:* Through integration with, for example, physical access control or central station automation software, the system can become even more powerful. Control doors, map movements, and handle alarms more efficiently. Add remote services for event handling through mobile devices.
- *Terms of contract:* The business model in hosted video differs as it is set up around a monthly service fee rather than an up-front CAPEX. The contract normally defines service levels and for how long the contract is binding. The service provider needs to lock the subscribers into an agreement so that they, over time, can earn back their up-front investment.



# CHAPTER 15

## Intelligent video

These days, massive amounts of video are recorded. However, the sheer volume of recordings and lack of time mean that much of the material is never watched or reviewed. As a result, events and incidents are overlooked, and suspicious behaviors are not noticed in time to prevent crime from happening.

Intelligent video, also known as analytics, video analytics, video content analysis, or intelligent video analysis, can bridge this gap. Analytics is a broader term that includes audio and nonvideo sensors such as passive infrared sensors. Analytics applications take video, audio, and other types of input, convert it to data, and analyze the data to find events of interest. For example, some applications recognize car license plates, while others focus on protecting critical infrastructure through virtual lines that can trigger alarms and alerts in the event of an intrusion. The development of new video analytics systems is ever growing, matching many types of security and efficiency needs.

Analytics-based systems are never idle. They scan video and audio in real-time 24/7, looking for information, events, or threats and responding immediately by initiating recordings or alerting security staff. Analytics can significantly reduce demands on network bandwidth and storage space, free up staff so they can do other tasks than constant monitoring of numerous cameras, and enable smart search to quickly pinpoint specific events.

Analytics-based systems can also extract data from surveillance video streams and integrate the information into other applications such as retail management or access control systems, creating new benefits and opening up a wide array of business opportunities.

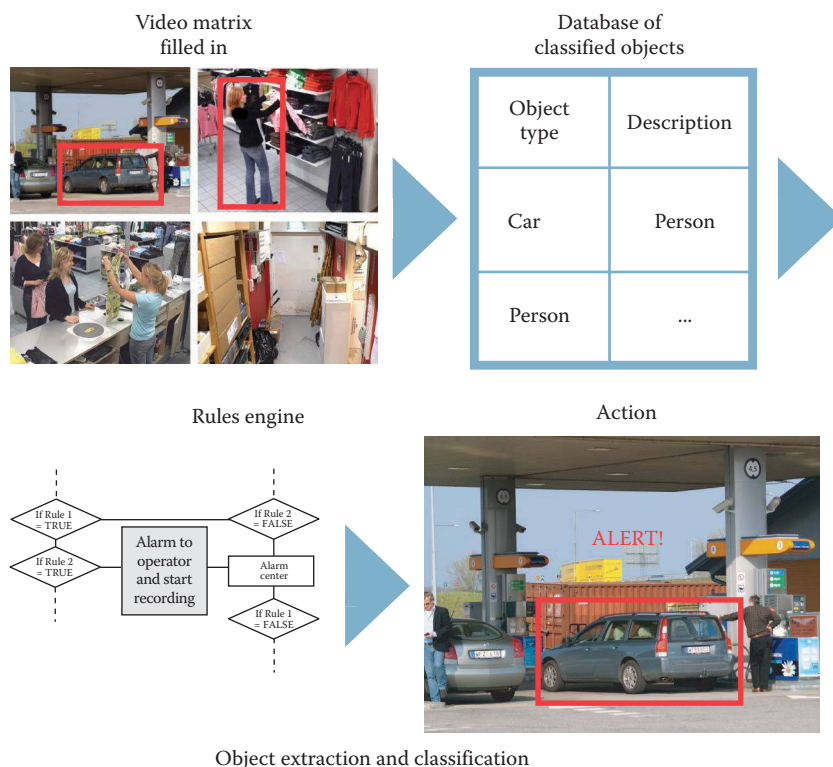
This chapter describes the basics and history of intelligent video, as well as the architectures and standards applicable to analytics. The next chapter discusses the most common analytics applications.

### 15.1 WHAT IS INTELLIGENT VIDEO?

---

Intelligent video or analytics is the process of analyzing video data with the goal of transforming it into actionable information. Analytics-based systems use complex mathematical algorithms to analyze the video and convert it into data. Typically, they extract moving objects or other recognizable forms while filtering out irrelevant movements (Figure 15.1).

The resulting data are stored in databases that can be searched with sets of rules applied, for example, an object passing a virtual line in the video or more than 10 cars waiting in a drive-through line. Rules can be programmed to help determine if the events observed in the video are normal or if they should be flagged as alerts to security staff or police.



**Figure 15.1** Intelligent video extracts actionable data from video images or streams.

Intelligent video is a vital component of critical security installations, supporting timely decision making in critical situations. Analytics applications, such as people or traffic counting, also open up new effective ways to manage businesses.

## 15.2 GENESIS OF INTELLIGENT VIDEO

In 1997, the Defense Advanced Research Projects Agency (DARPA) Information Systems Office in the United States began a 3-year program to develop video surveillance and monitoring (VSAM) technology. The objective of the VSAM project was to develop automated video and enable it to understand and evaluate the information received for use in battlefield surveillance applications. The technologies developed under this project enable a single human operator to monitor activities over a large area using intelligent video analysis. The analytics systems were designed to be autonomous, notifying the operator only if security threats occurred.

Many researchers at universities such as Carnegie Mellon University and Massachusetts Institute of Technology were chosen to develop a wide range of advanced surveillance techniques. They include the following:

- Real-time moving object detection and tracking from stationary and moving camera platforms
- Recognition of generic object classes (e.g., human, sedan, truck) and specific object types (such as a campus police car or courier van)
- Object pose estimation with respect to a geospatial site model
- Active camera control and multicamera cooperative tracking
- Human gait analysis
- Recognition of simple multiagent activities

- Real-time data dissemination
- Data logging
- Dynamic scene visualization

A lot of the intelligent video companies and technologies are spin-offs from the universities involved in the VSAM project.

## 15.3 WHY INTELLIGENT VIDEO?

The security industry is a growing and evolving industry. As video surveillance installations expand both in number and size, there is a demand for smarter software systems that enable management and security staff to tackle their surveillance challenges in a successful and effective way. In addition, as the security market shifted from proprietary, closed-analog closed-circuit television (CCTV) systems toward open and fully digital IP-based network video, new possibilities for harvesting non-security-related information from surveillance systems emerged and provided new user benefits. As the network video surveillance market matures, the expectations and demands evolve as well, resulting in a huge variety of analytics applications that are more sophisticated, more efficient, more specialized, and more cost efficient than ever before. Network video and analytics greatly simplify the process of integrating video streams with other IT and IS systems.

The three main market drivers for intelligent video can be summarized as:

1. Streamlining video surveillance operations
2. Managing stored video efficiently
3. Improving business operations

The following sections look into each of these three drivers.

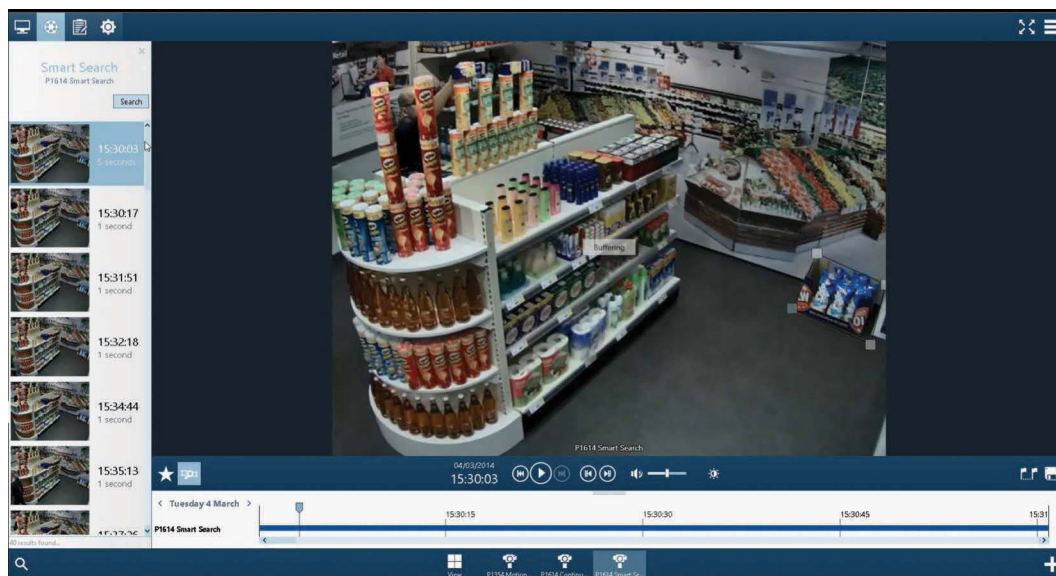
### 15.3.1 Streamlining video surveillance operations

The efficiency of the video surveillance in a typical security installation is limited because of one major issue: it is almost impossible to watch all the video all the time. For installations with a large number of cameras, it is obviously impossible for one or even several operators to watch all the cameras. Even in the unlikely scenario that there is one security guard for each camera, studies show that after only a short period of continuous video monitoring, operators often miss a large portion of the activity shown on the screen. After only about 20 minutes of uninterrupted viewing, test subjects overlooked almost all activities. If nobody is actively watching and acting upon the events shown on video, there is obviously an increased security risk to people and facilities.

Analytics presents a solution to this problem. Analytics applications analyze and filter the massive amount of information in multiple continuous video streams so that only relevant alerts are presented to security staff or police. With analytics, fewer security operators must spend their time monitoring video. Security managers can reduce their personnel count or invest the redundancy of hours in guards to patrol the premises and other preventive measures. Even very large video surveillance installations benefit from analytics because staff will not be expected to attentively watch dozens or even hundreds of monitors for hours to detect undesired activity or suspicious persons. Instead, an analytics-based system can do the job of alerting operators, for example, when people move into restricted areas, cars drive the wrong way, crowds gather, or someone tampers with a video surveillance camera.

### 15.3.2 Managing stored video effectively

Finding incidents in stored video would be extremely time consuming if the operator actually had to sit and watch the recorded video to find the footage. Even if the operator was experienced and could watch video at four times the speed of real time, large archives of video would still take a long time to search manually. Searching also implies that you know that an incident has occurred.



**Figure 15.2** After a smart search, the video management system shows thumbnails of the relevant clips based on parameters, such as area or time, as set by the operator.

The odds of finding an incident of which there is no physical proof are low. With smart search, operators can find the right video clip much easier (see Figure 15.2). For example, they can go to a camera view and draw a shape around the area of interest and also specify a time frame during which they suspect or know that there has been an activity. The video management system (VMS) finds and displays the relevant video clips.

A major retailer's study showed that only approximately 1% of its recorded video is ever watched. In some applications, the percentage is even less.

An even more significant issue is the cost of streaming and storing video. Continuous high-resolution recordings demand a lot of bandwidth and server space. Most of the time, there are no incidents to discover. Therefore, saving hours and hours of continuous recordings is a waste of time, storage, and money. Given these factors, most video files are simply stored and typically deleted after 7–30 days.

Analytics applications such as motion detection and face recognition ensure that only relevant video footage is recorded and stored. This minimizes the need for network bandwidth and storage space and reduces the amount of relevant video data that must be searched. In addition, some analytics-based systems can automatically search through days or even months of stored video to find security incidents in a matter of seconds.

### 15.3.3 Improving business operations

Analytics applications make it possible to use video for other things than security. Retail is one of the segments that is pushing the frontier of analytics. They use video intelligence for business intelligence purposes such as consumer behavior analysis. For example, through people counting, dwell time tracking, and heat mapping, users can find out how many people stopped at a particular merchandising shelf, which register is the most popular, what time of the day the customer traffic peaks, if cash is handled the right way, or which route customers take through the store. In this way, analytics makes it possible to extract greater benefits from a video surveillance installation, enabling a higher return on investment.

Big data is one of the megatrends in today's technology and business environment. Modern businesses are bombarded with huge amounts of data from different sources. The idea of big data is that there is a great opportunity for businesses to improve their operations if they can find effective

ways to manage and analyze all that data. Video content analysis, combined with point-of-sale data, loyalty card data, and the like, gives retailers a great opportunity to evaluate behaviors and improve operations.

## 15.4 INTELLIGENT VIDEO ARCHITECTURES

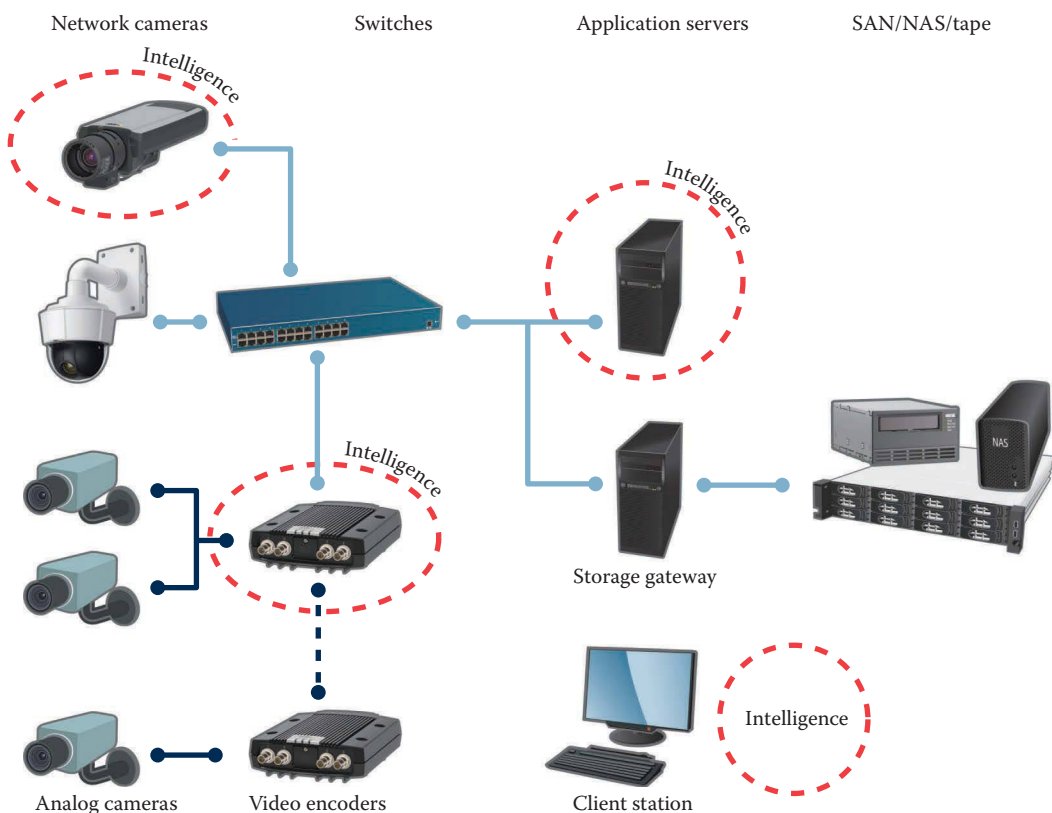
Intelligence can reside in many parts of a video surveillance system (see Figure 15.3). There are two broad categories of architecture for implementing analytics-based systems:

1. *Centralized architecture:* Video and other information are collected by cameras and sensors and brought back to a centralized server for analysis.
2. *Distributed architecture:* The network cameras or video encoders are smart devices capable of processing the video and extracting relevant information. This architecture is also referred to as intelligence at the edge.

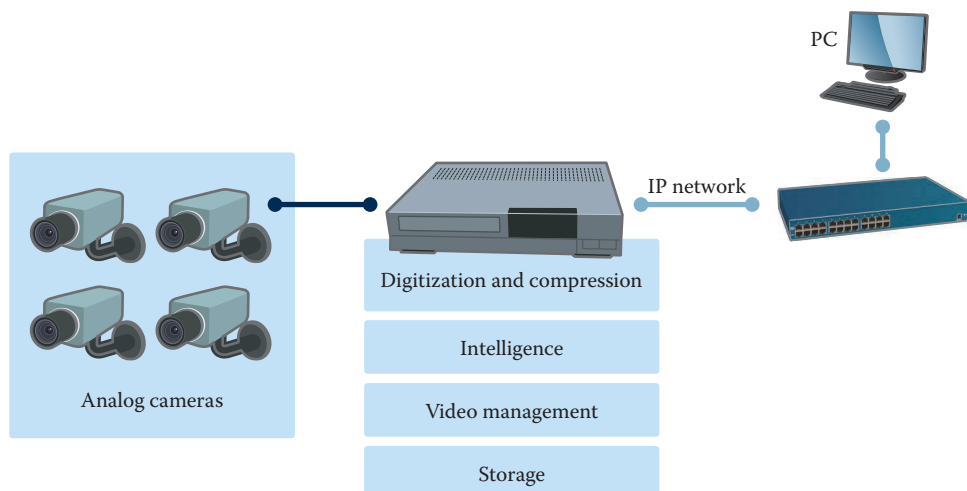
Through rational design and load distribution, an analytics-based system can improve performance and lower the overall costs substantially. The following sections look at centralized and decentralized systems and how different components of a video surveillance system can be used to implement intelligent video.

### 15.4.1 Centralized systems

In centralized architectures, all the video from the cameras is brought back to the headend for centralized processing. Legacy infrastructures with analog cameras mostly use traditional multifunction digital video recorders (DVRs) at the headend, whereas in a network video system, PC servers are used for video processing.



**Figure 15.3** The intelligence can be located in different parts of a video surveillance system, creating centralized or distributed architecture.



**Figure 15.4** In a digital video recorder (DVR)-based installation, the analytics functionality is located in the DVR, along with all other functions, such as digitization, compression, video management, and storage.

#### 15.4.1.1 DVR-based installations

When using traditional CCTV systems, the surveillance video from analog cameras feeds into an intelligent video-enabled DVR (Figure 15.4). DVRs have encoder cards that convert the video from analog to digital format and then perform the intelligent analysis (e.g., people counting or car license plate recognition). They also compress the video, record it, and distribute resulting alarms and video output to authorized operators. DVRs have limited computational power. Therefore, when they run analytics that require a lot of processing power, you need to cut back on the number of cameras.

In this architecture, coax cables connect each analog camera to a DVR. Most DVRs are embedded devices. Some use proprietary video formats. Although this approach works well enough for small installations with a limited number of cameras, it is not scalable or flexible.

#### 15.4.1.2 PC server-based installations

To overcome the limitations of DVRs, newer centralized architectures use commercial off-the-shelf PC servers for video processing (Figure 15.5). The video from network cameras is brought directly to servers over a network. If the cameras are analog, video encoders digitize the video first before transmitting it over a network to a server.

This architecture is more flexible and scalable than proprietary DVR-based architectures because digitization and compression have been pushed out to the network cameras and video encoders. However, because the servers perform many of the processor-intensive tasks (transcoding the video, managing the storage, and processing the video for analysis), they need considerable processing power. Therefore, each server is only able to process a relatively small number of cameras.

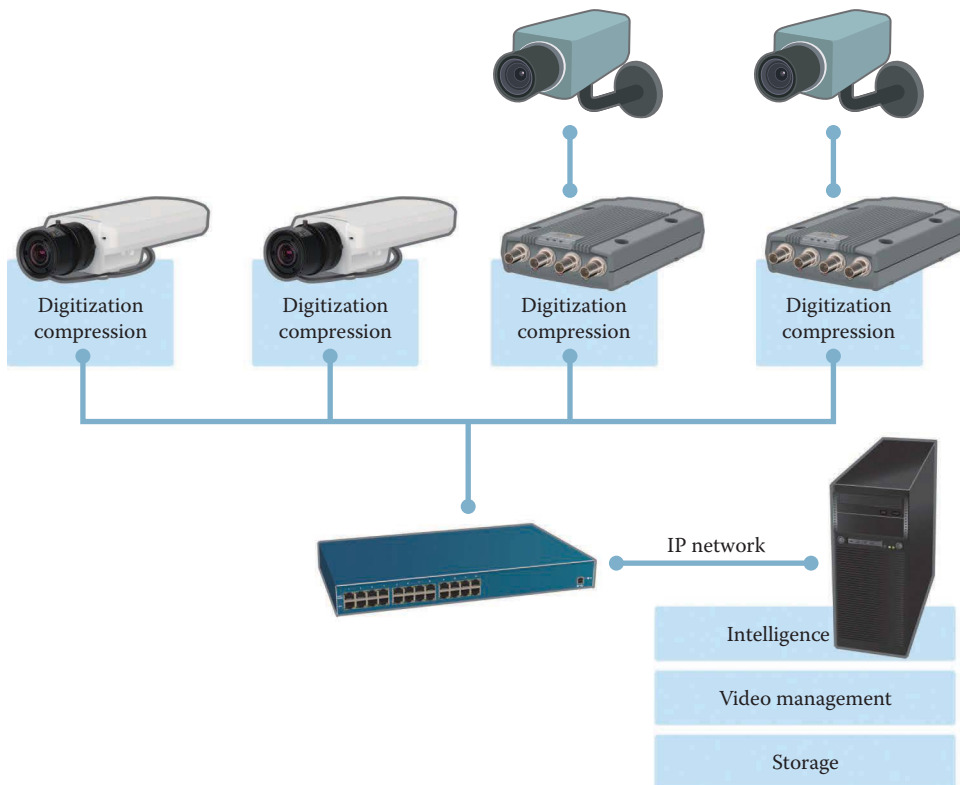
### 15.4.2 Distributed systems

Distributed architectures are designed to overcome the limitations of a centralized system, which tend to overload a central point such as a PC server or DVR. Distributing the processing to the edge in a network also reduces the bandwidth consumption.

#### 15.4.2.1 Intelligence-at-the-edge installations

The most scalable, cost-effective, and flexible architecture is based on “intelligence at the edge,” which means that the network cameras and encoders do most of the video analysis. (Analog cameras do not have the capability to analyze video.)





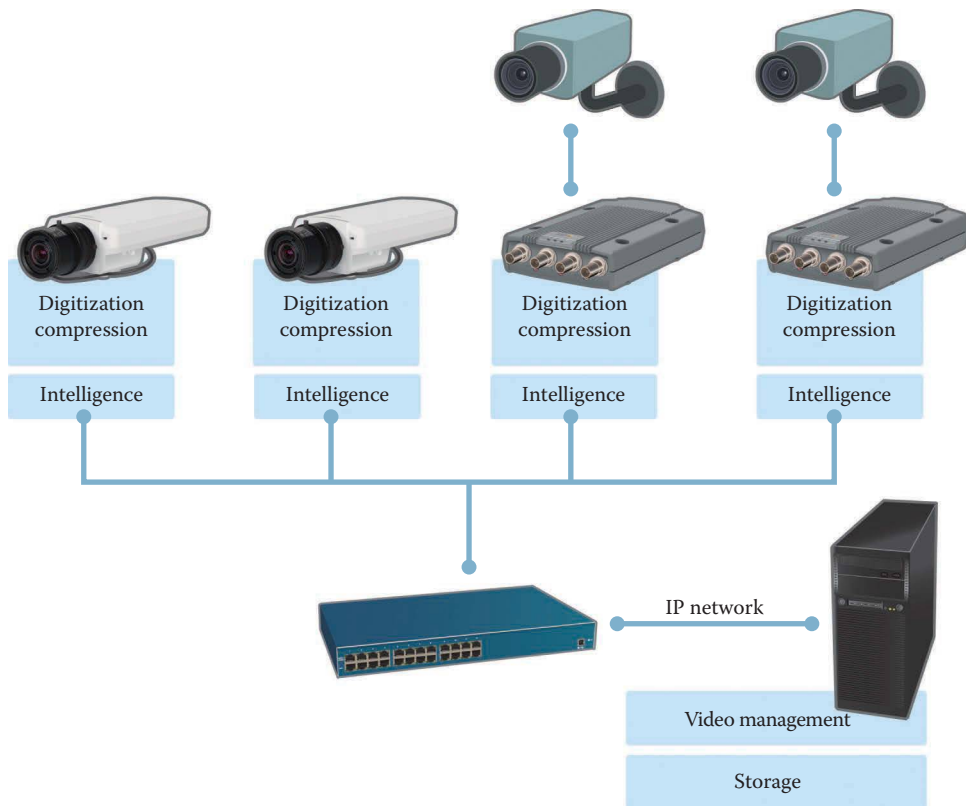
**Figure 15.5** In a PC server-based installation, the analytics functionality is located in the server, along with all other functions such as video management and storage.

Network cameras and video encoders with, for example, video motion detection can make use of this feature by sending video only when motion is detected in defined areas of a scene (Figure 15.6). If no motion is detected, no video is sent. The load on the infrastructure, including the required number of operators, drops dramatically. When using specialized applications such as license plate recognition or people counting, the impact of running applications in the camera is dramatic: the cameras can extract the required data (license plate information or number of people) and send only these data, perhaps with a few snapshots.

This type of architecture uses the least amount of bandwidth because the cameras can send out metadata and intelligently figure out which video to send. This significantly reduces the cost and complexity associated with a video analytics system and completely eliminates the drawbacks of a centralized architecture.

Another advantage of having video processing at the edge is that it significantly reduces the cost of the servers needed to run analytics applications. When the intelligence processing is done in the network cameras, the servers can handle many more video streams. In some applications, where just the data are needed and not the video (e.g., people counting or license plate recognition), the resulting data can be sent directly into a database, which further reduces the load on servers.

The quality of the video analysis is much better when it is performed in cameras with edge intelligence because the cameras can process raw video data before the information is reduced through video compression. Video-processing algorithms are more effective when processing raw video than compressed video. In centralized architectures, compressed video is sent to the servers because streaming raw video would take up too much bandwidth. To process compressed video packets, the



**Figure 15.6** A truly distributed system with intelligence at the edge (in the network cameras and video encoders) is the most scalable and cost effective.

servers have to decompress or transcode them. This requires more servers for each given number of cameras, which increases the cost.

The cost-saving benefits of architecture based on intelligent edge devices can be summarized as follows:

- *Fewer PC servers are needed for video processing:* Fewer servers also mean lower power consumption and less maintenance costs. And in certain environments, such as retail stores, where there are generally no server rooms, installing a large number of servers is simply impractical.
- *Lower demands on bandwidth:* Streaming only essential information means lower data rates and that lower-cost network components can be used.
- *Higher accuracy:* When doing video analytics on the edge, the video available to analyze is uncompressed and hence includes more details than compressed video, giving higher accuracy of the analysis.

With intelligence at the edge comes the opportunity to create a self-contained recording system, where the video management software is hosted on a PC, on a network-attached storage (NAS), or on the camera itself. Users control and manage the video through web browsers or apps on mobile devices. These systems are completely edge based and are not dependent on network and cable infrastructures. They are also perfect in environments that have limited network connectivity and no on-site server resources.

Edge intelligence is essential in cloud-based systems (hosted video). It is especially cost effective in applications such as people counting, where the metadata are more important than the video. Cameras and encoders with edge analytics process the video and send only the metadata and a low-bandwidth

stream to the cloud. Edge storage (NAS or onboard storage) can be used to store a stream with higher resolution and higher frame rate. To read more about hosted video, see Chapter 14.

### 15.4.3 Integrating intelligent video applications

Intelligent video has become popular over the last few years. Increased accuracy and improved ease of use has driven popularity and affordability, along with a desire to make the most out of new and existing video surveillance systems.

Many manufacturers of video surveillance equipment and providers of video management software include analytics functionalities with their products. These intelligent video functionalities—for example, video motion detection—are suitable for most installations. Occasionally, other analytics functionalities for specific applications are provided, such as license plate recognition and people counting.

Building robust and commercially viable analytics applications with a high level of functionality requires a great deal of expertise in specific sciences. For example, image analysis and biometrics use advanced mathematical algorithms. In addition, applications often require specialized knowledge in a certain field, such as retail, public transportation, or customs control. For this reason, a number of software companies focus their skills on developing and supplying analytics applications that solve specific needs in specific markets, allowing for tailored video surveillance solutions.

Although this progress brings greater freedom of choice for the user, it also makes it necessary for manufacturers to ensure compatibility and ease of integration between network cameras, video management software, and analytics applications. To be commercially attractive, all cameras, software, and analytics applications must be easy to use. They must also build on open, published platforms and application programming interfaces (APIs). This enables easy installation of plug-in camera software (see Figure 15.7) and communication with VMSs.

**Application Settings: IPConfigure Embedded LPR**

**Application Settings**

[Main page](#)

[Embedded LPR User Interface](#)

Status: Running

**Parameter Settings**

Font set: USA (VA)

Target framerate: 9 fps

Plate color: Dark text on light background

Minimum character confidence: 50 [0..100]

Minimum plate confidence: 60 [0..100]

Overview frame count: 3 [..5]

Flip raw input: ☐

Disable LED: ☐

FIFO processing: ☐

Push results: All detected plates

Push method: HTTP POST

Push format: JSON-encoded plain text

Push address: http://192.168.101

Log level: 5

Save Reset

**Figure 15.7** An example of plug-in software for license plate recognition. (Image courtesy of IPConfigure, Norfolk, VA.)

From the analytics suppliers' point of view, using open platforms and standards provides the most flexibility and makes the software easier to sell. It is this openness that allows vendors to install analytics applications from different manufacturers in their cameras and systems.

## 15.5 APIs AND STANDARDS

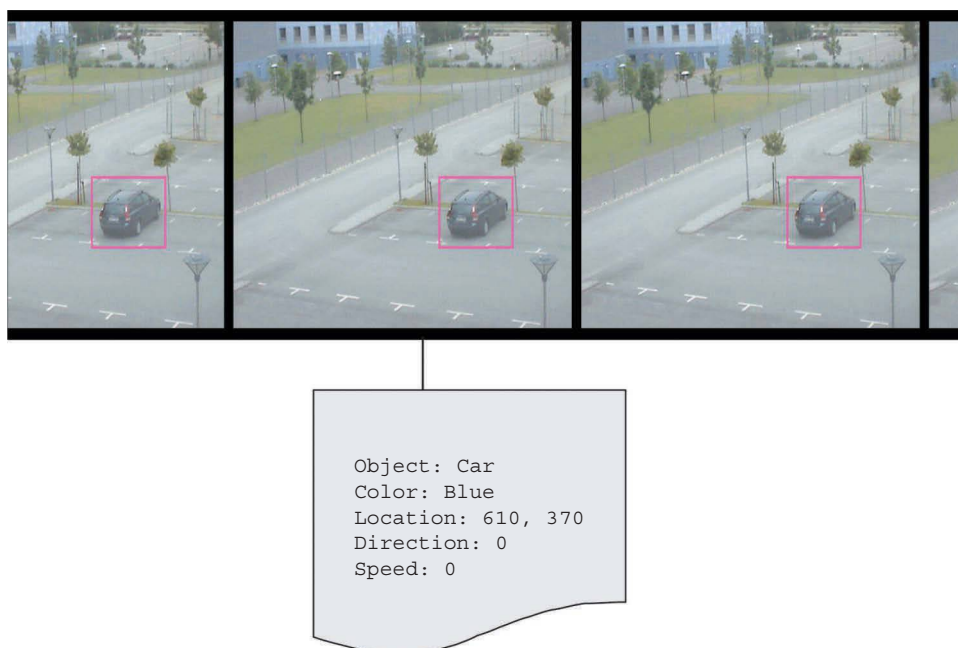
Analytics applications work on digital streams of video that come either from network cameras or from video encoders, which digitize analog video from analog cameras. For intelligent video applications to work as part of a video surveillance system, there must be standardized formats for digital video streams. A number of video compression standards exist, some of which are more relevant for intelligent video than others. These include Motion JPEG, MPEG-4, and MPEG-4 AVC/H.264 and HEVC/H.265. For more information about compression, see Chapter 6.

In addition to the requirement of having video compression standards, there is a need for open standards to describe and tag the content of video images. Information about the content of an image (or any content at all) is called metadata, which is explained in more detail in the succeeding paragraphs.

### 15.5.1 Metadata

Metadata, which literally means “data about data,” provides a solution to the challenge of sifting through volumes of recorded video to find, filter, and retrieve the right information efficiently. Metadata allows for tagging images with information about the images. The metadata tags are what enable automatic analysis of video streams and users to quickly find exactly what they are looking for in a recording.

For example, an analytics application that counts people passing through an area can tag a number to an image as part of the process of streaming it to a central server for further analysis. The analytics application can be designed to send only the tag information—the number of people—and not the stream images at all.



**Figure 15.8** A video stream tagged with metadata—for example, the object, color, and location of a car parked in a parking lot.

For video surveillance of a bridge or a highway, an analytics application could tag vehicles appearing in an image according to certain criteria and store the tags together with the video. This would make it possible for a user to search based on these criteria and instantly have access to the right video sequences. For example, the operator can search for all blue cars heading north in the past 24 hours, instead of manually searching or watching 24 hours of recorded video (Figure 15.8).

When the metadata functionality is based on open and standardized ways of describing information, it can easily integrate with various other systems. It allows for scalable and open systems. A number of standards exist for attaching metadata to data streams, for example, Extensible Markup Language (XML) and Scalable Vector Graphics (SVG), and many have been developed for specific areas such as libraries, geospatial use, and database management.

## 15.6 BEST PRACTICES

---

The market for intelligent video is growing. There is a wide variety of offerings from many companies. Although the architecture of the solution is often overlooked, it is important to understand it fully, or you might end up with the wrong solution. When choosing to implement analytics, the solution should be based on open standards and be able to scale as an installation grows.

Things to consider include the following:

- *Reliability of the system:* Is the architecture minimizing the risk of system failure and the downtime associated with it?
- *Scalability and flexibility:* Can the system scale effortlessly from a few to many cameras? Can it intelligently distribute processing on different components of the network?
- *Interoperability:* Are the analytics applications and video management applications tied to specific hardware? Is it possible to combine system components from different vendors?
- *Data format used by the analytics application:* Is the metadata based on an openly published standard or API so that it can be easily incorporated into other systems?
- *Accuracy:* No video analytics system is 100% accurate, although as regards impact, the difference between 90% and 95% accuracy is large. An operator can only deal with a few false positives a day.
- *Integration:* Is it a stand-alone system that operates without or beside a VMS? Is the system integrated and tied into a central recording platform? Is the system centralized or distributed? Depending on the system, different opportunities and choices arise.
- *Built-in or plug-in analytics:* Are the analytics built into the camera or is the camera platform open? Open platforms allow users to choose among the best of breed of plug-in applications.



# Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>



# CHAPTER 16

## Intelligent video applications

Today, there are a large variety of intelligent video applications (also known as video intelligence, video content analysis (VCA), video analytics, or just analytics) to choose from, and as the market continues to grow, so will the development of new and smarter analytics software. Some analytics tools, such as video motion detection (VMD) and camera tampering alarms, are relevant to most video surveillance installations. Others are specially developed to address the needs of specific industries, such as retail, transportation, and critical infrastructure markets. This chapter provides an overview of the most common types of analytics applications.

Video analytics was a hot new technology in the security industry in the early 2000s. Some unfulfilled promises and early challenges gave the technology a black eye for many years. Since then, analytics has restored its reputation. It is used often, it is used by many, and its benefits are recognized. It is still important to have realistic expectations, as TV shows like *CSI* still make some users expect more than the technology can deliver. This is discussed at the end of this chapter, along with some best practices.

### 16.1 CATEGORIZING VIDEO ANALYTICS

---

The following sections further discuss the different types of analytics technologies and their uses, but first an introduction about the different ways to categorize analytics.

#### 16.1.1 Categorizing video analytics by technology

A seemingly straightforward approach to categorizing intelligent video applications is to describe them based on the technologies they use. Typically, video analytics are divided into the following technology categories:

- *Pixel-based analytics*: Used to send alerts when loss of video quality or detection of motion in images occurs. Basic VMD and camera tampering detection are common examples of pixel-based analytics.
- *Object-based analytics*: Based on the recognition and categorization of objects in an image. This is the category with the largest number of applications, and it is further divided into two broad subcategories: object tracking and object recognition and classification. People counting and advanced video motion detections are examples of object-based analytics.
- *Specialized analytics*: These applications use both pixel- and object-based intelligence to process video for a specific application. Number and license plate recognition (LPR), facial recognition, and fire and smoke detection are examples of specialized analytics.

### 16.1.1.1 Pixels, blobs, and objects

At a basic level, video analytics software analyzes every pixel in every frame of video, characterizes those pixels, and then makes decisions based on those characteristics. Basic analytics applications (such as pixel-based motion detection) make decisions based on changes in the characteristics of pixels. When a certain number of pixels change based on criteria such as size, color, and brightness, they can raise an alert or trigger other actions.

Blob recognition involves a level of intelligence beyond detection of pixel changes. A blob is essentially a collection of contiguous pixels that share particular characteristics, and blobs have boundaries that delineate them from other parts of a video frame. Blobs can be analyzed and characterized as being particular objects. For example, a blob can be identified as a person or a car by analyzing its shape, size, speed, or other parameters. Applications based on object classification and tracking require the most sophisticated software algorithms (Figure 16.1).

### 16.1.2 Categorizing video analytics by use

Many intelligent video applications are designed to extract different kinds of information from video, to process information in different ways, and to apply different rules for making decisions. As described earlier, intelligent video applications can be categorized based on the technologies behind them, but what really matters in the end is how they can be used to make life easier and safer for organizations and businesses.

The analytics applications described in this chapter are categorized based on how they are typically used:

- *Analytics for security:* Primarily used in safety management applications and often in real time. Event-based analytics use video data to improve surveillance system efficiency and to facilitate investigations. The camera triggers real-time events and alarms. An operator analyzes the alarm and decides on an action.
- *Analytics for business intelligence and operations:* Primarily used for business management and evaluation purposes and not for immediate action. For example, what are the traffic patterns in the store, do people visit a promotional corner, or are people loitering in restricted area?
- *Analytics for hybrid operations:* Used by both real-time security operations and business intelligence. For example, LPR or facial recognition can be used for real-time purposes as well as reporting purposes. LPR applications with white-and-black lists can be used to open gates when white-listed cars approach (an event-based operation), but also to create parking invoices (a reporting operation).



**Figure 16.1** Examples of detection and tracking of (a) pixels, (b) blobs, and (c) objects.

## 16.2 ANALYTICS FOR SECURITY

To users, analytics for security act as detectors that answer yes-or-no questions: Is something there? Did something move? Did something new happen? Did someone tamper with the camera? Whenever the system detects an activity, it can trigger alerts to operators or automatic system responses.

Examples of analytics for security are:

- VMD
- Camera tampering detection
- Object tracking
- Fire and smoke detection

### 16.2.1 Video motion detection

VMD is the original and the most basic and widespread video analytics application in the surveillance industry. It is primarily used to reduce the amount of video stored. By only storing video in which changes occur, security personnel can store video for a longer period at a given storage capacity.

So how does the VMD application know which video to record and which to ignore? The application flags video clips that have changes in them (motion) and ignores the video clips that have no changes (static). VMD can also flag events, for example, when persons enter restricted areas, to operators for immediate action.

VMD is the foundation for a large number of more advanced intelligent video applications, such as people counting, digital fences, and object tracking. Worth noting is that VMD is very different from detecting motion with a PIR sensor. A PIR sensor detects heat by measuring the infrared light emitted from objects and people in its field of view. For more information about infrared technologies, see Chapters 4 and 5.

#### 16.2.1.1 Evolution of VMD

Algorithms continually compare images from a video stream to detect changes in an image. The first generation of applications that recognized motion in the camera view did so simply by detecting pixel changes from one frame of video to the next. Although this scheme certainly helped reduce the amount of storage required (by not storing video in which nothing changes), it was not very useful for real-time applications because it gave too many false positives. It raised too many alarms based on pixel changes caused by events of no interest (such as minor light changes, slight camera motion, or motion of trees).

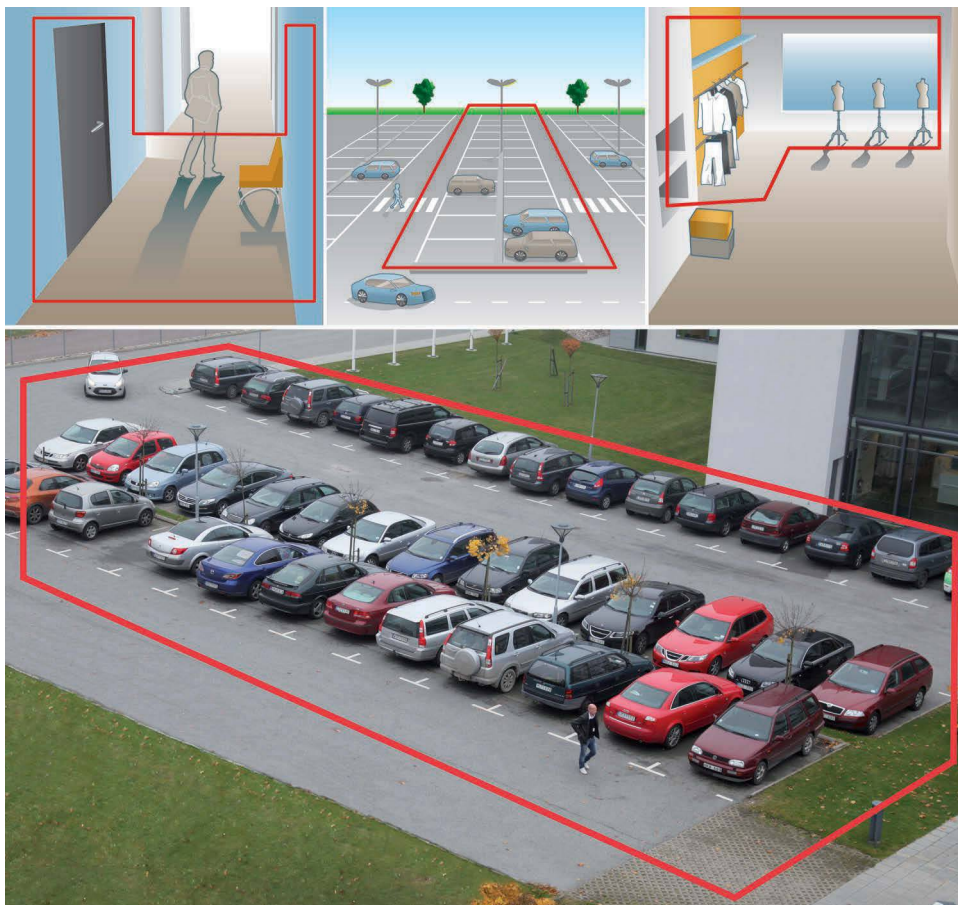
More advanced VMD systems can exclude pixel-based changes from known sources, such as natural changes in lighting conditions based on the time of day or other known and repetitive changes in the camera's field of view. This drastically decreases the number of false alarms.

Recent generations of VMD applications are more sophisticated and deploy algorithms that are more advanced. They not only detect individual changes in pixels but also group pixels together or detect motion on an object level. This way, the system can understand that many pixels together actually constitute larger objects such as people or cars. This decreases the number of false alarms even further.

#### 16.2.1.2 Tuning of VMD parameters

VMD applications usually have controls for tuning the detection parameters. Examples of parameters are:

- Ignore filters for different object types
- Object-size thresholds
- Lingering times



**Figure 16.2** Video motion detection applications are especially suitable for low-traffic areas such as office corridors, parking lots, and unattended shop areas. In the interface, the user draws polygons around the areas that are detection areas. Within the detection area (include area), the user can draw exclude areas around objects that should be ignored by the application. When objects, such as people or vehicles, move into or within the include area, the application detects their movement and can trigger other actions, such as sending an alarm to an operator, turning on floodlights, starting recordings, or activating a speaker or siren.

The balance between different settings directly affects the number of false alarms and whether all relevant motion in the scene is detected. For best results, fine-tune the VMD after installing the cameras. To ensure a robust implementation, keep a log for a while and observe how the settings affect the VMD events in each camera.

Advanced network cameras (Figure 16.2) often allow more exact placement of detection areas. It is also possible to create multiple detection areas, where each area can use different parameters.

Whenever there is an incident, operators and investigators usually find it helpful to see what happened before and after the main event. Many network cameras and encoders keep prealarm buffers and postalarm buffers that, for example, can start 30 seconds before the event and continue 15 seconds after the event. This ensures that when the VMD detects motion, the system always has video of what happened in the seconds before the event, during the event, and the seconds after the event.

### 16.2.2 Camera tampering detection

Imagine a video surveillance installation in a subway with thousands of cameras. One of the cameras has been mistakenly redirected. For weeks, it has been pointing in the wrong direction. Then one day,

there is an incident, and investigators need to review video from that location. However, the video from the camera in question is useless because no one has noticed that the view has changed.

Moving or covering cameras or their parts can be discovered by camera tampering detection. Because it is the obstruction or view change that the application detects, it does not matter to the system whether the interference was intentional or not. For example, the lens of a camera could be deliberately or accidentally covered by paint, powder, moisture, a sticker, or other material (see Figures 16.3 and 16.4). The cameras could be deliberately or accidentally redirected to a view of no interest. More serious cases of tampering can include interference with the focus or actual removal of cameras. Without camera tampering detection, these or similar threats would go undetected and the recorded video would be useless.

Camera tampering detection is applicable in any installation but is used predominantly in environments that are potentially exposed to vandalism, such as schools, prisons, or public transportation.

A tampering detection application must be capable of telling the difference between predictable changes in a camera view versus unexpected changes caused by tampering. Otherwise, false alarms would counteract the benefits.



**Figure 16.3** Camera tampering detects in seconds when a camera is repositioned, goes out of focus, or is covered.

### Camera Tampering

**Tampering Period**

Minimum duration:  seconds ▾

**Dark Images**

☒ Alarm for dark images.

**Figure 16.4** Some camera tampering detection will learn the scene automatically, making setup very easy. Simply select if camera tampering detection should be activated, along with the minimum duration, and the camera takes care of the rest.



As with many analytics applications, there are different ways of implementing tampering detection. It can be implemented in a camera, or centrally in software or on a server. By installing tampering detection algorithms in each camera, the system scales more easily than when run through a central server.

### 16.2.3 Object tracking

In most security operations, it is necessary to keep track of where objects and people are, inside a facility or around the perimeter of it. Intelligent video applications based on object tracking process images in a specific way. First they segment a particular object in a camera view, and then they track that object as it moves around in the view or as it moves from one camera to another.

VMD used to be pixel based, but modern applications are mostly object based instead. Object tracking is better than pixel-based detection at removing false positives in outdoor environments, such as flags, bushes, and other swaying objects; rabbits, birds, and other small animals; and sweeping headlights, reflections, and other short-lived objects.

#### 16.2.3.1 Crossline detection

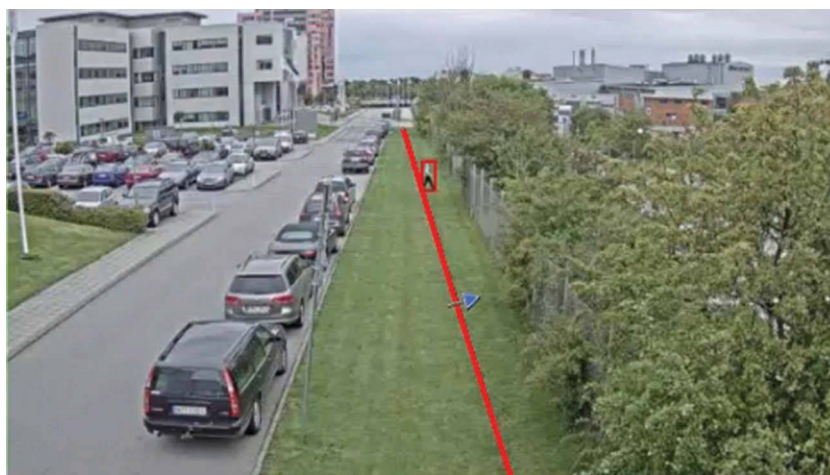
Crossline detection applications, also known as tripwire, perimeter guard, fence guard, digital fence, or virtual fence, are used to alert security personnel to possible perimeter breaches. Setting up such alerts usually involves designating a line or area and then telling the system to generate an alert if objects pass that line in a particular direction or if objects enter or leave a certain area (Figure 16.5).

An example might be someone trying to enter through an exit door at an airport terminal. Another example might be setting up a system to allow staff to *leave* a building after dark but to send an alarm if someone tries to *enter*.

#### 16.2.3.2 Intrusion detection

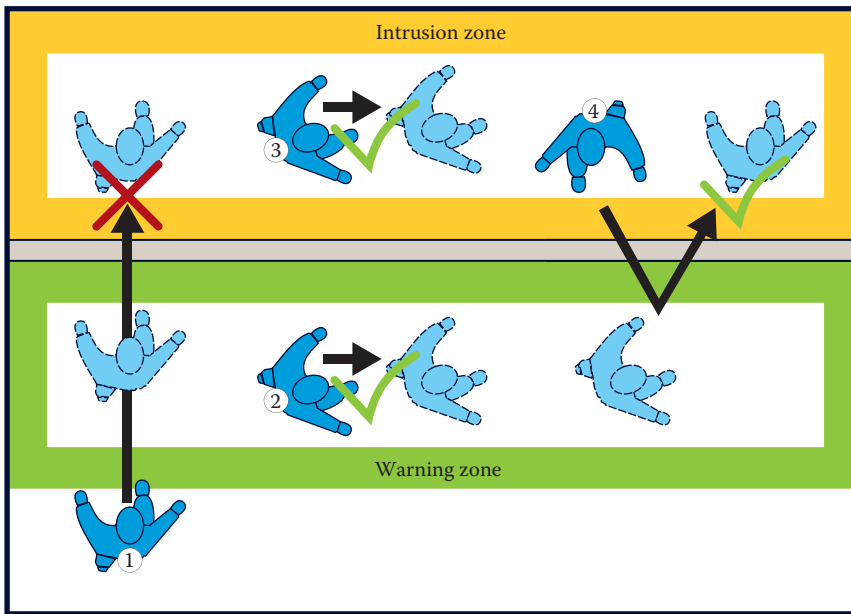
Intrusion detection is fundamental to the protection of critical infrastructure, such as transportation systems, oil and gas plants, power and water treatment plants, police and military agencies, and hospitals. What needs to be protected is typically a fenced area, and what it needs protection from is unauthorized people entering, vandalizing, or moving in the forbidden zones. Advanced intrusion detectors adapt to the perspective in the video image and use trajectory algorithms. This means that they can detect which direction people are moving in and where they came from, even if they pass behind things like trees and pillars.

It is possible to set up multiple detection zones with different rules for allowing movement, zone crossing, sending alarms, and recording video. For example, anyone is allowed to move within the warning



**Figure 16.5** Crossline detection can be used to detect and send an alert if someone passes the boundary between an open and a gated area.





**Figure 16.6** Foes may not cross to an intrusion zone (1), but they may move within a warning zone (2). Friends may move within the intrusion zone (3) and also move from an intrusion zone to a warning zone and back again (4).

zone, but no one is allowed to cross from the warning zone to the intrusion zone. People may also move within the intrusion zones. Because they originate from the intrusion zone, the system identifies them as “friends,” but if they originally came from a warning zone, then the system identifies them as “foes.” If a foe tries to cross to an intrusion zone, the system sends an alarm to the operator. See Figure 16.6. Typically, the system starts recording whenever it detects movement in any of its detection zones.

#### 16.2.3.3 Object left behind

Object left behind is often a critical application for the security of common areas. The application targets threats from explosives left behind in bags or packages. It watches an area and keeps track of all objects in it. When an object that was previously moving becomes stationary and stays that way for a certain period of time, the system raises an alert and shows the operator the object of concern (Figure 16.7).

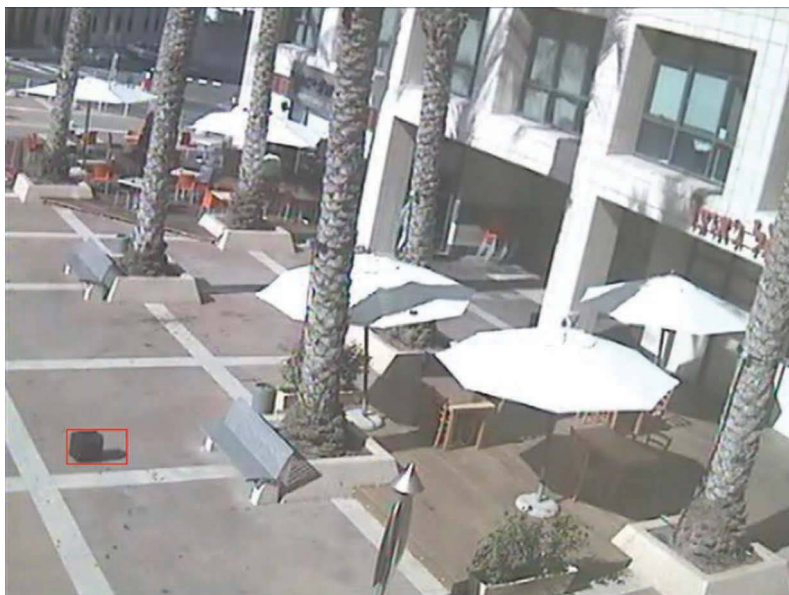
#### 16.2.3.4 Loitering detection

Loitering functionality tracks the amount of time and the number of people who linger in a certain area. For example, in a parking lot or in front of a bank ATM, lingering or loitering people could indicate malicious intent. Dwell time software, which is used for business intelligence, analyzes similar patterns (Figure 16.8); see Section 16.3.3.

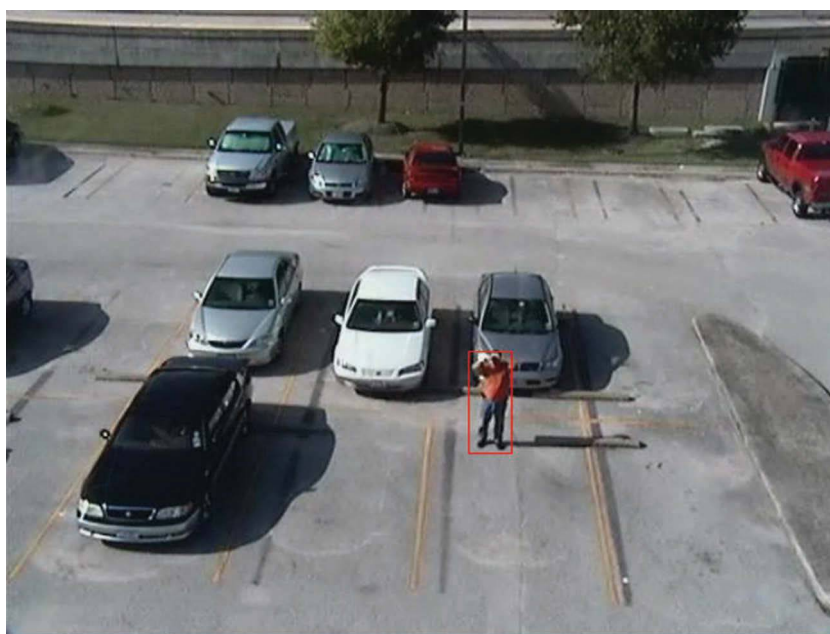
### 16.2.4 Fire and smoke detection

Traditional smoke detectors trigger when there is a certain amount of smoke in a room. Fire and smoke detection as part of analytics applications process video streams and look for visual cues to fire or smoke. This gives the potential advantage of being able to react faster than a smoke detector. An analytics-based fire and smoke detector may be able to react as soon as flames are visible in a room, rather than later when the smoke has reached the ceiling. Such an application can be a good complement to traditional types of smoke detectors (Figure 16.9).

The intelligent video system processes video images and reacts when combinations of color, light, and movement typical for fire or smoke appear. Then, an alarm can be sent together with relevant video images to an alarm monitoring central.



**Figure 16.7** An example of a system for abandoned object detection. (Image courtesy of Agent Vi™, Rosh Ha'ayin, Israel.)



**Figure 16.8** An example of a system for loitering detection. (Image courtesy of Agent Vi™, Rosh Ha'ayin, Israel.)

## 16.3 ANALYTICS FOR BUSINESS INTELLIGENCE AND OPERATIONS

Video analytics for business intelligence and operations is becoming a popular way to get more value out of video surveillance systems, especially in retail and transportation environments. The software produces reports with information about, for example, how people move in the area, what their patterns are, or if they are displaying unusual or unwanted behavior.



**Figure 16.9** Fire and smoke analytics help protect lives and property. They are essential in high-risk environments, such as power and gas stations (a), flour and paper mills (b), warehouses (c), and waste disposal sites that house large amounts of flammable material (d). (Image courtesy of Araani, Kortrijk, Belgium.)

Analytics for business intelligence and operations deliver many types of information, help analyze business strategies, and make studying customer behavior easier.

Analytics for business intelligence and operations includes:

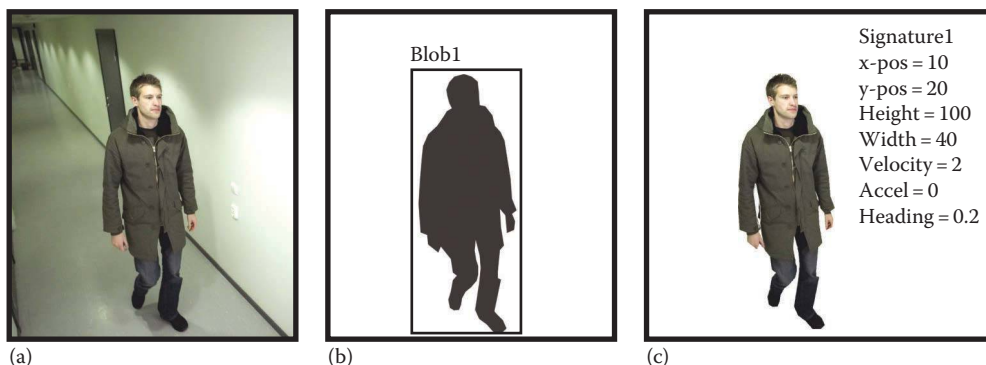
- Object classification
- Object counting
- People counting
- Dwell time and heat mapping
- Traffic management

### 16.3.1 Object classification

Most analytics software generally goes through the following steps (Figure 16.10):

1. *Detection* analyzes all the pixels in video frames, compares pixels in each frame to a reference frame, and figures out which objects are moving.
2. *Segmentation* extracts the moving objects and assigns descriptive signatures to them, that is, descriptions based on criteria such as color, size, direction, and time.
3. *Classification* sorts the segmented objects into different object types, such as a person or car. Then, they are assigned a set of descriptors that characterizes them, for example, color, size, or direction.

The ability of video analytics-based systems to recognize object types and isolate the object of interest greatly enhances their accuracy and usefulness. Does the application count all objects in a

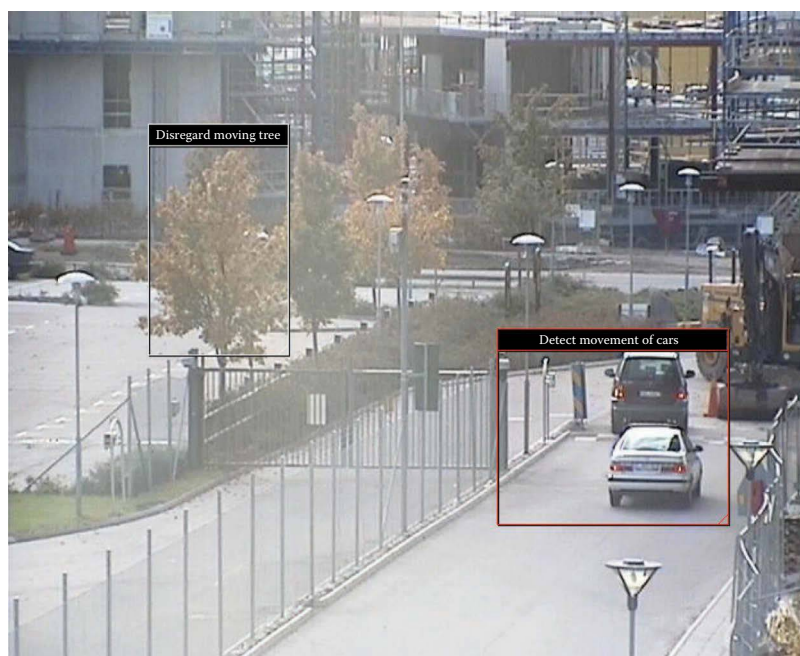


**Figure 16.10** Object-based analytics software detects and segments a moving object first (a), classifies it (b), and provides a set of metadata that describes the object (c).

scene, or can it recognize and count only particular objects, such as people or vehicles? (see Figure 16.11). Is the program able to tell the two people in a handholding couple apart, or would they be interpreted as one object?

A specific challenge for analytics applications is that objects can appear in other configurations than expected. For example, the system may be able to distinguish a human being from a dog. This is based partially on the knowledge that humans have a different aspect ratio; that is, they are substantially taller than they are wide. However, if a person is crawling, their proportions differ substantially from the norm. To be reliable, a video intelligence system needs to compensate for these types of deviations from the norm and be able to recognize a human being regardless of whether the person is crawling, crouching, standing, or running. This ability is often called aspect ratio compensation.

The classifiers and characteristics of the objects in an image are called metadata. Once metadata is collected, it can be compared to a set of criteria for action, such as a person walking the wrong



**Figure 16.11** Video analytic systems need to distinguish between object types so that they only call attention to objects that matter to the user.



way, a bag left behind, or a car entering a restricted area. If the criteria are fulfilled, the system can raise an alert in real time or retrieve the matching video from storage. For more information about metadata, see Section 15.5.1.

### 16.3.2 Object and people counting

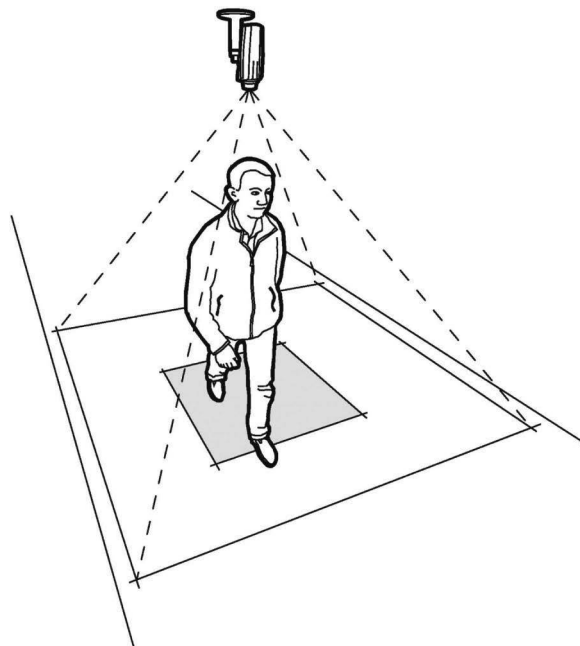
After detecting and classifying an object, the intelligent video system can use the information for different purposes. One purpose is to count the number of objects that behave in a certain way, such as a person walking in the wrong direction. Another application can be to determine the total number of objects of a particular type in an area.

When using people-counting software, the placement of the camera becomes important (Figure 16.12). Ideally, the camera should be placed immediately above the entrance. The exact height depends on the camera lens and the width of the entrance. Any person who passes under the camera must be larger than 6% of the camera's total horizontal field of view. The camera must produce images with good enough quality to distinguish the people passing under the camera. Although other technologies (such as infrared) can also count people, video-based people counting can, in many cases, provide better accuracy.

People counting is useful in retail stores or other environments where it is important to know the number of people entering or exiting an area. The data are used to understand customer behavior; to better plan product placement, promotions, and advertisements; and ultimately to increase the return on investment.

The ability of video intelligence systems to accurately count the number of people is at the core of a range of applications, including:

- Customer traffic monitoring
- Queue management
- Tailgating



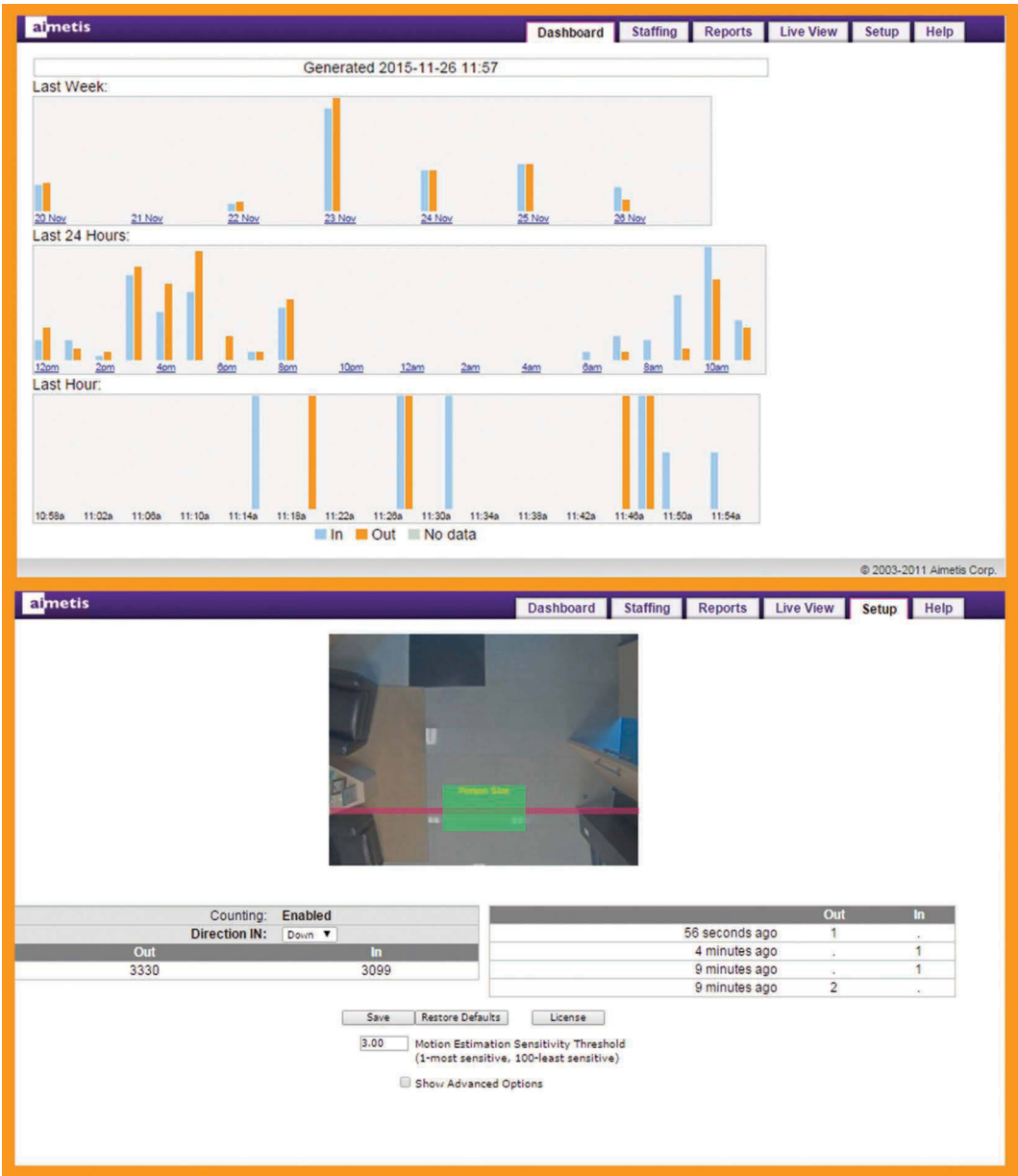
**Figure 16.12** The placement of the camera becomes important in providing the appropriate accuracy.

16.3.2.1 Customer traffic monitoring

A retail store can use people counting to count the number of people who enter and exit the store, go through certain aisles, or stop by a particular merchandising display (Figures 16.13 and 16.14). By comparing the traffic count with their point-of-sale data, store managers can calculate their conversion rate. A franchise or chain can use the data to compare the performance between different stores.

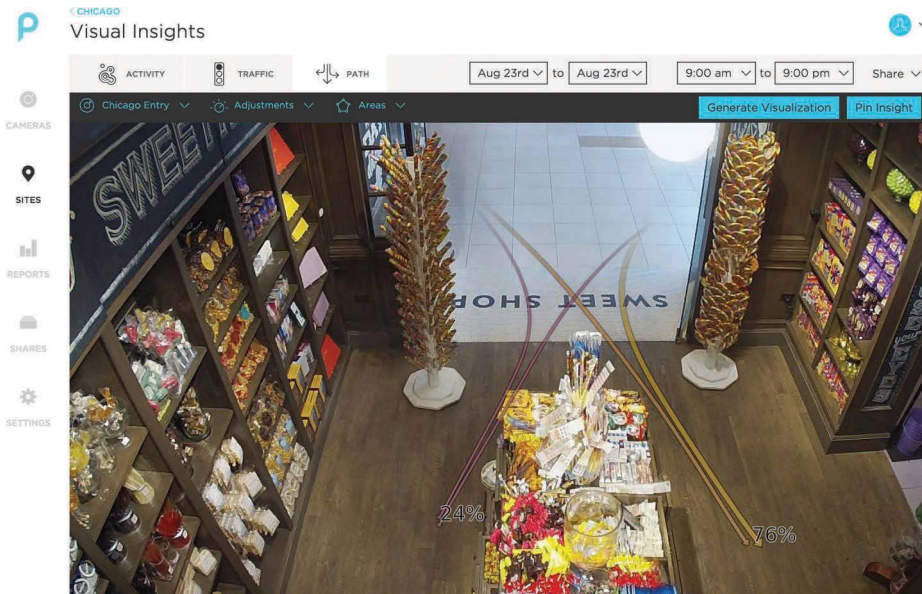
16.3.2.2 Queue management

Also known as staff optimization, queue management software counts the number of people standing in various queues, for example, retail checkout counters, customer service desks, airport

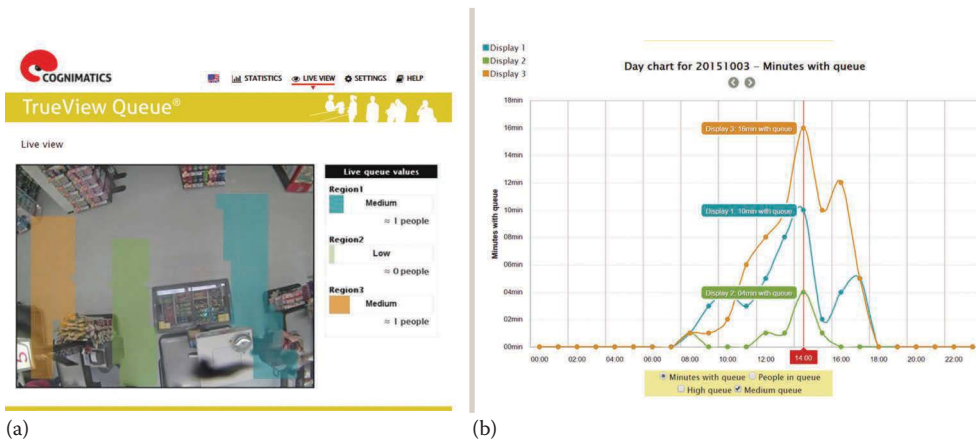


**Figure 16.13** An example of using people counter software to keep statistics of customer traffic. (Image courtesy of Aimetis, Waterloo, ON, Canada.)





**Figure 16.14** An example of customer traffic monitoring. The application shows which direction most visitors choose when they enter the store. (Image courtesy of Prism Skylabs, San Francisco, CA.)

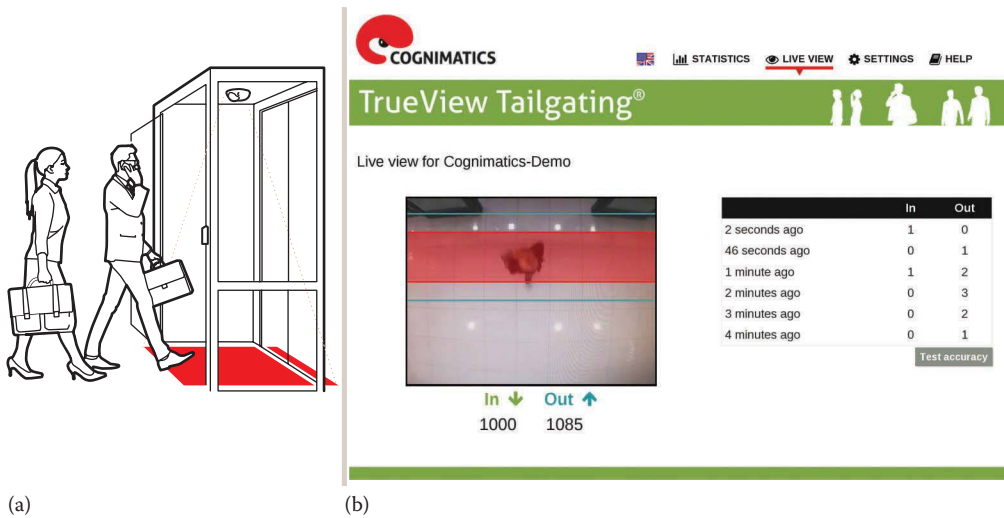


**Figure 16.15** An example of queue management software with live view queue monitoring (a) and queue statistics (b). (Image courtesy of Cognimatics, Lund, Sweden.)

check-in counters, and passport and security controls (Figure 16.15). The information is often used to improve customer service. For example, when a queue passes a specific threshold, the software can prompt the staff to open more checkouts. The software also keeps statistics over time so that managers can improve their understanding of customer traffic across different seasons, get indicators of how good the staff response is, and make smarter decisions when planning work shifts.

### 16.3.2.3 Tailgating

People counting can be used with access control systems to make them more secure. The main principle of access control is that only people who belong in a building should be able to enter its doors. People-counting software can send an alert when multiple people enter a facility although only one person has swiped their badge (Figure 16.16).



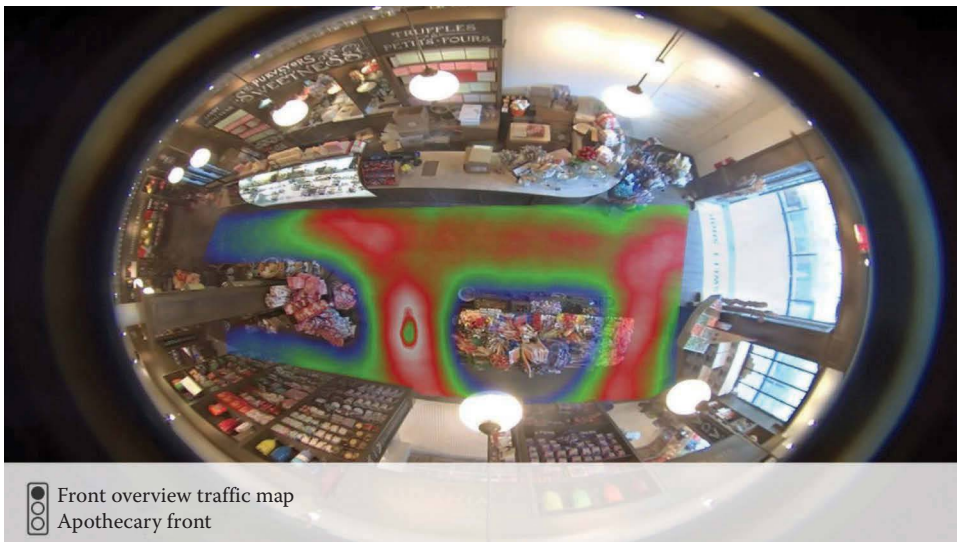
**Figure 16.16** An illustration of tailgating (a) and example of software (TrueView Tailgating®) for tailgating detection (b). (Image courtesy of Cognimatics, Lund, Sweden.)

### 16.3.3 Dwell time and heat mapping

Similar to loitering detection, dwell time software keeps track of how long people stay in an area and which are the typical traffic paths. However, rather than preventing unwanted behavior like loitering detection does, dwell time focuses on how customers connect with products and promotions, how many people pass by a specific display or area, what their level of engagement is, and how many of them actually end up purchasing the exhibited product (see Figure 16.17).



**Figure 16.17** An example of using dwell time software to support merchandising and advertising decisions. (Image courtesy of Cognimatics, Lund, Sweden.)



**Figure 16.18** An example of a heat map, showing the areas of high traffic and low traffic. (Image courtesy of Prism Skylabs, San Francisco, CA.)

Heat maps, also known as shopper activity maps, show customer traffic patterns in a very visual way (see Figure 16.18). They put colorful overlays over the video stream, showing which zones are hot and which are cold. Their color palettes look very similar to those of thermal imaging, where blue is cold and red is hot. But of course the analyzing algorithm is based on object classification and people counting, not on infrared radiation.

Dwell times and heat maps provide data that help when optimizing the layout of a retail store. For example, the software can give visual proof if an endcap (i.e., a product display at the end of an aisle) is not attracting the expected amount of interest. The manager can then compare it to areas that have the opposite effect and change the displays accordingly. Just as you want to steer the traffic to specific displays, you want to avoid bottlenecks that frustrate customers and deter sales. It is all about creating the best flow and getting the most out of the space, customers, and staff.

### 16.3.4 Traffic management

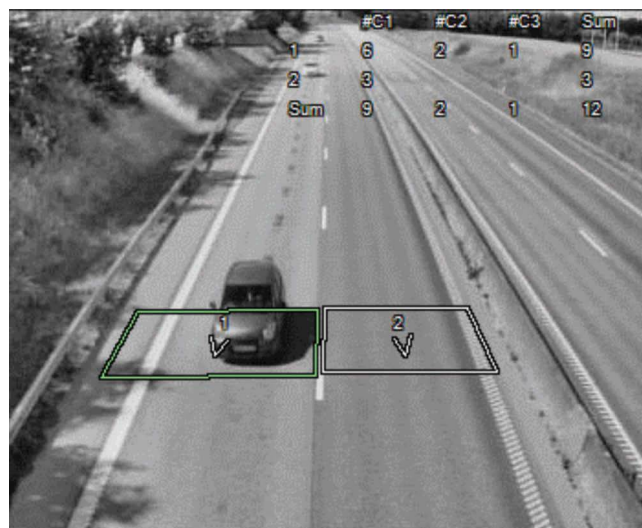
Traffic management software is used to monitor and improve the overall traffic flow of urban areas. It efficiently replaces induction loops and other types of magnetic sensors to detect and count vehicles and pedestrians. Beyond the obvious benefit of not having to dig up the road to improve the system, traffic management software gives both statistics and video that show the status of the traffic in real time. How many cars pass through each lane? What is the length and speed of each car? In which direction are they moving? What is the traffic density of lanes and intersections? How long are the lines and what is the average waiting time? Of course, traffic management analytics can make a distinction between cars, heavy trucks, and motorbikes. Some also detect pedestrians and smaller vehicles such as bicycles.

Typically, you use multiple streams to serve different purposes—one low-quality stream in black and white for analytics (see Figure 16.19), one medium-quality stream in color for monitoring, and one high-quality stream in color for recording.

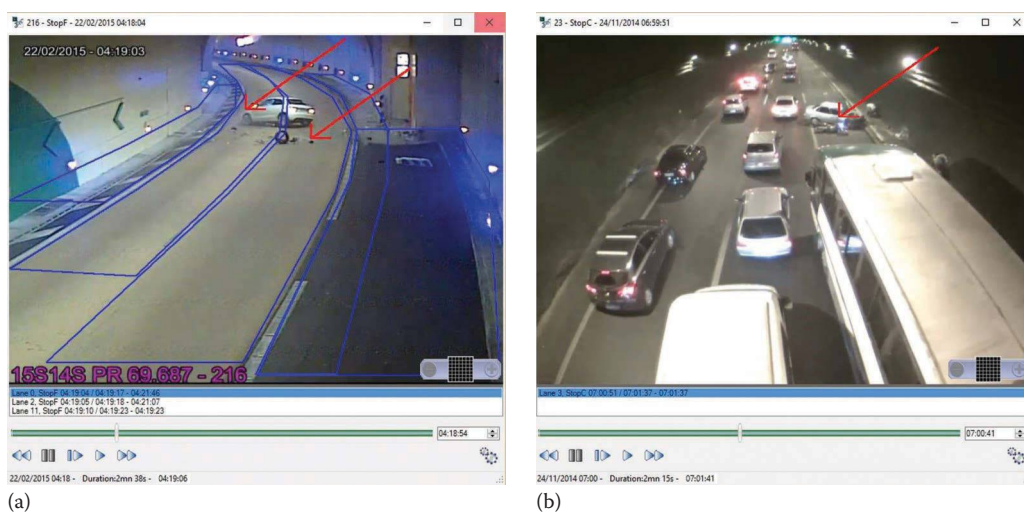
#### 16.3.4.1 Incident detection

With automatic incident detection (AID), emergency agencies and traffic controllers can get real-time alerts whenever traffic is hindered in any way. For example, AID software detects when vehicles drive too slowly, stop for too long, or stop in a forbidden area, when the road is congested, when a driver is steering their car oddly or in the wrong way, and when there are debris or pedestrians





**Figure 16.19** An example of traffic analytics. The focus is on keeping track of vehicles, collecting data, and detecting incidents. For this type of analytic, the system can use a lower-quality stream. If an incident does occur, the system starts recording a high-quality stream so that events can be analyzed thoroughly. (Software by Citilog.)



**Figure 16.20** Examples of incident detection (a, b). The arrows help the operator detect what has happened. (Images courtesy of Citilog, Arcueil, France.)

on the road. Traffic controllers get visual and audible alerts so they quickly can take adequate measures, such as alerting ambulances and warning other road users to slow down. This minimizes the risk of secondary accidents and helps prevent injuries and deaths, as well as infrastructure damage. Figure 16.20 shows examples of incident detection.

## 16.4 HYBRID ANALYTICS

Hybrid analytics applications have multiple uses. Security officers, crime investigators, and business managers can use them both for real-time event monitoring and for reporting purposes. For example, border control officers can use LPR to detect wanted vehicles trying to cross a border, and police can use it to investigate which vehicles were involved in a crime. Facial recognition can be used to open doors for people with access to a building or to identify the perpetrator of a street robbery.

### 16.4.1 Autotracking

Autotracking, also known as digital autotracking, is an analytics application that can pinpoint and follow individuals within a camera's view. Either the operator selects a person to track or the system automatically detects and tracks a person. The former is useful in crowded locations such as at airports and retail stores. The latter is used mainly for perimeter control, where even the presence of a single person is of interest (Figure 16.21).

Autotracking applications work best with cameras with a wide field of view, which allows an operator to track people over a wider area. For example, some network cameras have a 180° field of view that allows zooming into areas without losing video quality. Very-wide-view cameras, so-called 360° panoramic cameras, are also well suited for this type of tracking, provided they have high enough resolution to provide the required detail. If the objective is identification, then a mechanical pan, tilt, and zoom (PTZ) camera with optical zoom is probably a better option, as it gives much better pixel density (see Section 16.4.2).

### 16.4.2 Autotracking using PTZ

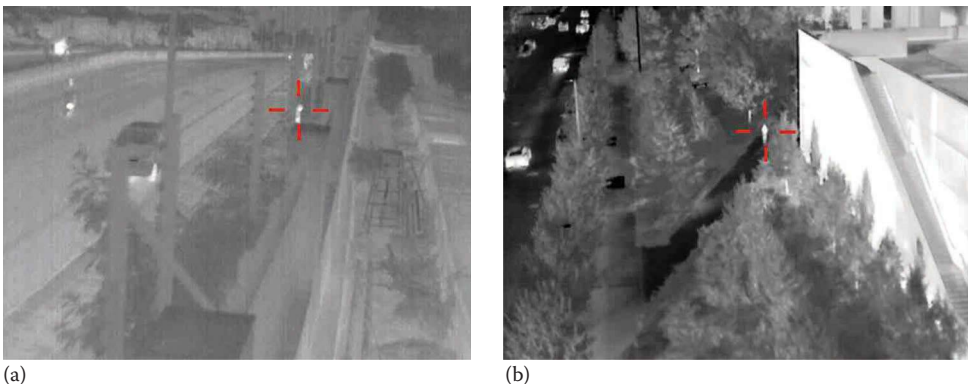
PTZ autotracking is an intelligence that automatically controls a PTZ camera to keep a person in sight. The advantage is that the camera can zoom in automatically to give a better view of the person. If only one PTZ camera covers the area, the disadvantage is that it might be pointing in the wrong direction at the time of a particular event. A hybrid approach would remedy that issue. Fixed cameras would find the events, and a PTZ camera would do the people tracking. Another issue with PTZ autotracking is that only one object can be tracked at a time, and some PTZ tracking systems get confused if more than one object is in the camera view.

There are two different types of automatic tracking:

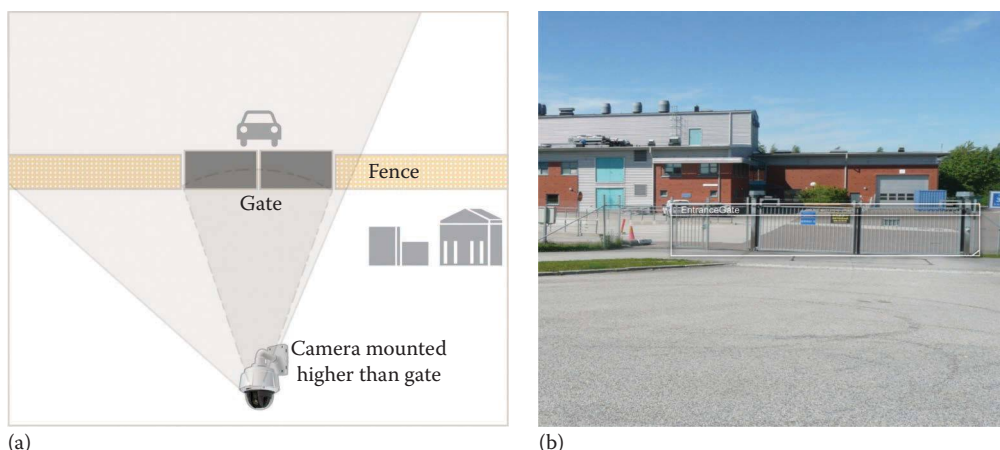
- *Automatic selection—automatic tracking:* The camera locks onto the first moving object until it loses that object. The camera will then find another moving object. This solution is useful in low-activity environments such as parking lots and hallways. It provides a view of the object without the need for on-site security staff.
- *Manual selection—automatic tracking:* The surveillance officer selects an object to track, and the camera follows it. This setup helps the officer focus on the object instead of being distracted by operating the camera.

#### 16.4.2.1 Gatekeeper

A PTZ camera with gatekeeper functionality can react to movement and automatically start panning, tilting, and zooming. Typically, the gatekeeper is used to monitor a gate or a specific area.



**Figure 16.21** A people-tracking system shows the current location of a person and tracks their movements. These images show people tracking with thermal cameras (a, b). (Images courtesy of Jemez Technology, LLC, Los Alamos, NM.)



**Figure 16.22** An illustration (a) and a screenshot (b) of the gatekeeper functionality. When an object enters the gated area, the gatekeeper zooms in and follows it.

When objects or people move in the gated area, the gatekeeper can follow them as they move away, zoom in on them, or move the camera lens to a preset position (see Figure 16.22). The gatekeeper can also trigger other actions such as reading a license plate (see Section 16.4.3) or recording video, which can be in the same camera, in other cameras, or in the VMS.

### 16.4.3 License plate recognition

LPR, sometimes referred to as automatic LPR (ALPR), has a variety of uses, ranging from access control to watching out for particular vehicles. For example, in an access control application, only vehicles with particular number plates are allowed access to a facility. In a criminal investigation, LPR can automatically look for certain sets of license plate characters to find suspect vehicles.

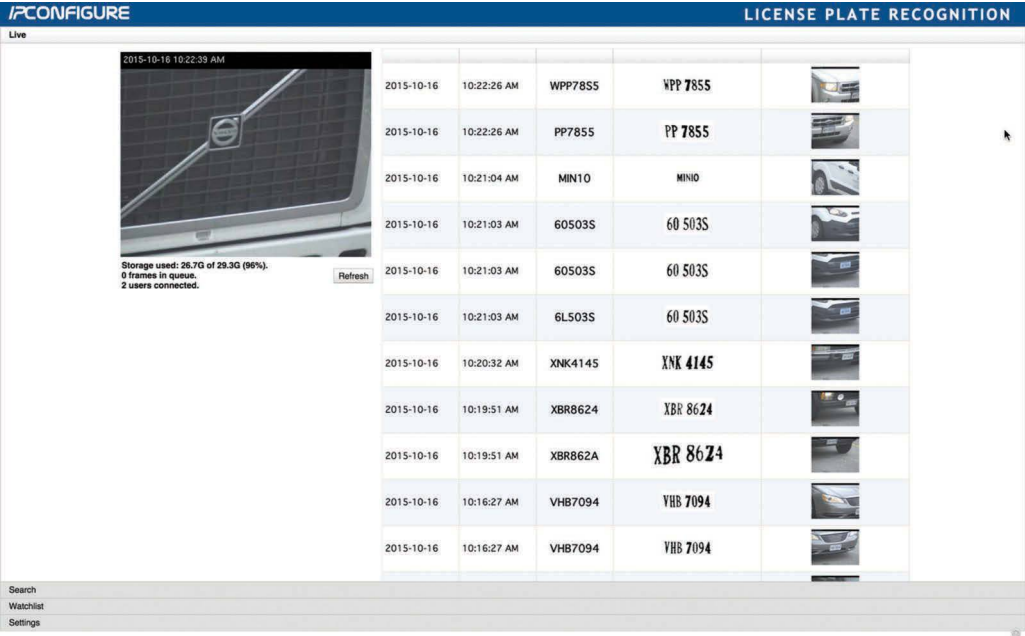
In a parking lot or parking house, LPR can automatically track the vehicles that enter, eliminating the need for a more expensive parking ticket infrastructure. LPR can also automatically monitor how long a particular vehicle stays parked. For example, a store might want to reserve its parking lot for its customers or to get an alert when a vehicle seems to have been abandoned.

In retail environments, LPR can identify cars whose drivers often shop in certain stores. This information can be used, for example, to analyze where these shoppers come from or to design direct marketing programs that reach only consumers in the correct geographical area (Figure 16.23).

An intelligent video application that performs LPR applies a process that consists of several steps. In the car identification example, the steps include the following:

1. *Find the car:* If the car is not parked in a known, well-defined physical location (such as the gate to a parking house), the first step consists of finding the car in an image. For cars in motion, VMD is a key element in making this work. For parked cars, it is a matter of recognizing the outline of a car in an image.
2. *Isolate the number plate:* Once the car is found, the next step is to isolate the actual license plate from the rest of the image. This is based on algorithms that define what a license plate looks like and where it might be mounted on a vehicle.
3. *Extract the characters:* The next step consists of extracting the letters and the numbers from the license plate using image analysis.
4. *Identify the characters:* Optical character recognition (OCR) transforms the characters from a collection of pixels into a stream of letters and numbers.
5. *Process the characters:* The final step consists of processing the resulting string of letters and numbers by storing it in a database or comparing it with existing entries.





**Figure 16.23** An example of a license plate recognition application. (Image courtesy of IPConfigure, Norfolk, VA.)

Some challenges are especially difficult for LPR to deal with. Bad weather, blinding headlights, and dirty or bent license plates can affect the result of the process. License plates also look different in different parts of the world, which means that the analytics application should be adapted to local conditions and quite often fine-tuned to the specific implementation. To provide the best images possible for the LPR application, specialty cameras are often used (Figure 16.24).



**Figure 16.24** In a license plate recognition (LPR) application, specialty cameras are often used to ensure that a clear capture of the number plate can be provided at all times of the day. If the camera is mounted on a car, it is possible to drive by parked cars in a parking lot and use the LPR application to find particular cars. (Image courtesy of Genetec™, Montreal, QC, Canada.)

From an implementation design point of view, LPR is a type of video analytics where the benefits of a distributed approach are very clear. When deploying an LPR application in a network camera, you can limit the transmission of data to the letters and numbers of a license plate—with perhaps just a snapshot of the vehicle—which drastically reduces the network load compared with a centralized implementation of an LPR system.

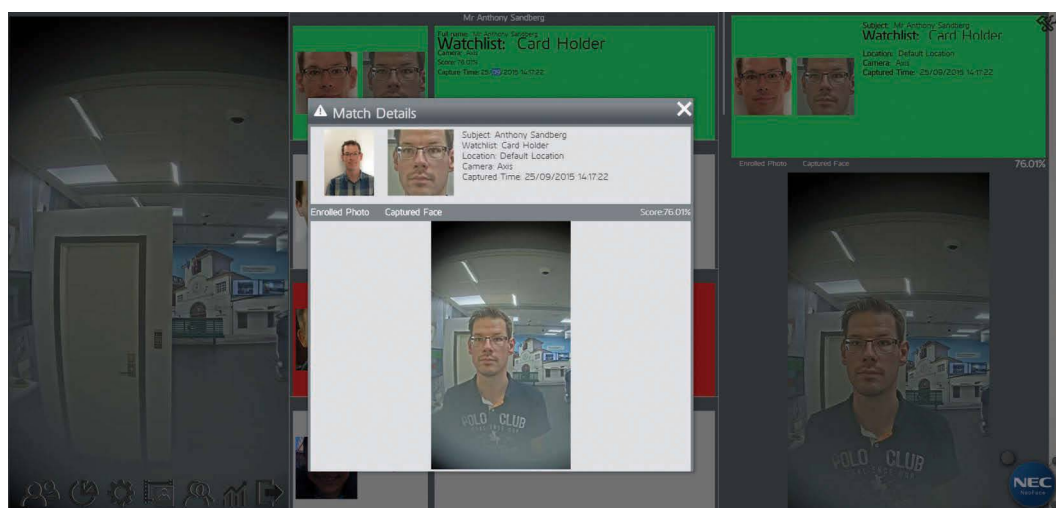
### 16.4.4 Facial recognition

Facial recognition has gotten a lot of attention, and its uses are many and varying. Police want to get alerts when certain individuals are seen in public or sensitive areas. Allowing only certain people to enter specific areas can enhance access control. For forensic purposes, the need can arise when searching for individuals in recorded videos. In casinos, managers may want to catch blacklisted individuals. Customs can use automatic searches to find certain individuals and improve the precision of border and passport control.

The facial recognition process is similar to the LPR system described earlier. However, one difference is that it is possible to define beforehand what the system should look for in a number plate—that is, a string of letters and digits grouped in a certain order. To know what to look for, facial recognition depends on the initial step of actually building a database of known faces. Examples of such databases are passport photo databases and police registers. Sometimes, getting those data becomes the biggest challenge in a facial recognition system.

Beyond that, the facial recognition steps include the following:

1. *Find the person:* If the person in question is not standing still in a predefined location, the first step consists of finding the person in an image.
2. *Find the face:* The next step is often called face finding, which means isolating the face from the rest of the body.
3. *Isolate features:* The subsequent step involves identifying typical parts of the face, in terms of locating the position and recognizing the shape of features such as the eyes, nose, mouth, chin, skin color, and hair. From these traits, a unique pattern of the individual's face is constructed. In some applications, face finding is sufficient. For example, in an airport, it can be used to measure the queue time from entering and exiting a check-in point. In this case, the actual identity of the individual is not of interest; the system merely should be able to separate one individual from another. Retail businesses can use it to calculate the conversion ratio; that is, how many customers, out of the total number of people



**Figure 16.25** An example of a facial recognition system. (Image courtesy of NEC NeoFace® Watch, Tokyo, Japan.)

who entered the store, made a purchase. Here, the system needs to distinguish individual faces, but their identity is completely irrelevant. In a monitoring situation, it may be enough for a guard to be presented with individual faces for matching against registered faces from access control badges.

4. *Match faces*: The final step involves matching the extracted face with signatures from a database to actually recognize individuals and make positive identifications (Figure 16.25).

As for LPR, the challenges for facial recognition systems are substantial. Even if lighting conditions are perfect, people generally move around and sometimes block each other, and faces change and age over time. Appearances can also easily be changed by a pair of glasses, a change of hair color or length, or the growing of a beard, or even through surgery. A particular challenge is the fact that people rarely look straight into a camera unless they are prompted to do so. To deal with this problem, developers have built 3D facial recognition software, which extracts 3D information from video streams and match it with a database.

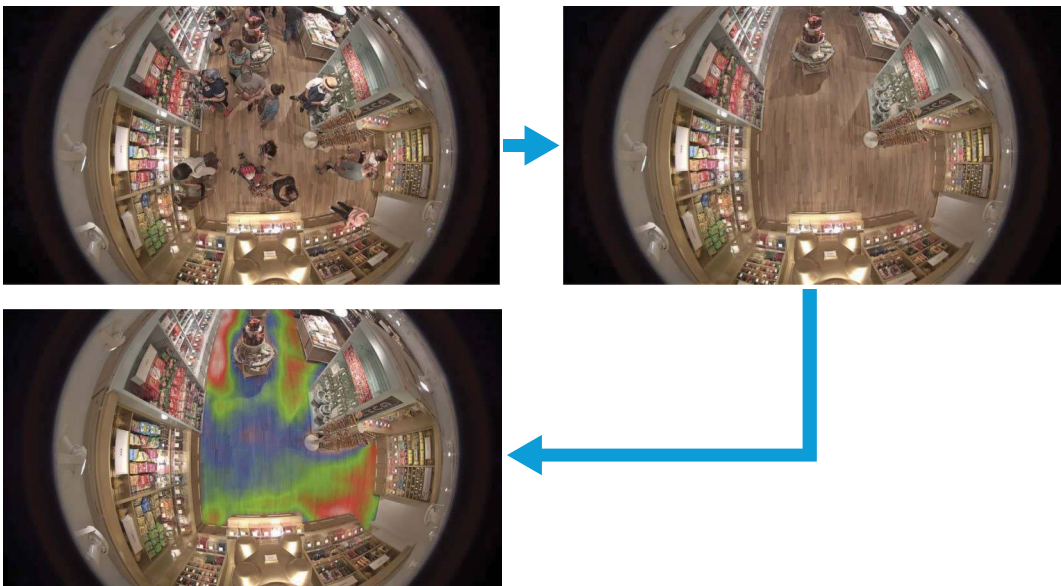
## 16.5 VIDEO ANALYTICS AND PRIVACY

Analytics applications, especially those that focus on identifying human features, are sometimes seen as being invasive of people's privacy. However, those concerns can be overcome by not storing actual pictures or videos of faces. Once data are collected, for example, from a people-counting application, there is no reason to keep pictures or video of individuals.

In some ways, video analytics is able to actually enhance privacy. For example, an application may be able to find and mask out all the people in a surveillance recording (see Figure 16.26).

### 16.5.1 Sound identification

As with video motion detection, audio detection is going through a generation shift. Many cameras have built-in audio detectors that can detect if a sound reaches a particular level. However, more advanced algorithms can imitate the human ear and make a distinction between different sounds. Because it uses temporal and spectral information (the relationships between connection, orientation, and distance) to analyze audio, the database of training samples can be quite limited. There are also fewer issues with low signal-to-noise ratios. Sound analytics are better than video



**Figure 16.26** A heat mapping application removes the people from the image and instead shows a heat map (also known as shopper activity map) of how people move and linger in different areas. (Image courtesy of Prism Skylabs, San Francisco, CA.)

analytics at providing early warnings of aggressive situations, gunfire, breaking of glass, or car alarms. When the application detects a sound, based on the defined parameters, it can trigger a camera to start recording and send an alarm to a security officer.

## 16.6 REALISTIC EXPECTATIONS ON VIDEO ANALYTICS

This chapter presented an overview of a variety of video analytics applications, from the more generic, such as video motion detection, to the more advanced systems, such as facial recognition.

Although all the applications described earlier can indeed perform well in a controlled environment, many will struggle to be robust and efficient enough to function optimally in real-life situations. Some systems will only work after tremendous tailoring efforts to meet every possible challenge in every situation.

Caught up in the enthusiasm of the potentials of analytics, the surveillance market initially had extremely high expectations on vendors and software suppliers to fulfill that potential. In retrospect, these expectations were too high compared to what is actually possible or cost effective. Although the market has calmed down in this regard, the industry still struggles with unrealistic expectations in many areas.

Generally, the analytics applications that enhance traditional camera functionality—such as video motion detection, image enhancement, and camera tampering—have matured well and can now be used in most applications without major tuning or modifications.

Specialized applications such as people counting, LPR, and crossline detection are also used more widely, although typically they require a fair amount of calibration in each installation. Some of the most advanced analytics applications should still be considered demonstration software rather than usable in the real world for the majority of users.

## 16.7 BEST PRACTICES

On a practical level, the key to success is to have a clearly defined use case. That is, what should trigger an alarm and what should not trigger an alarm are known and are possible to describe by simple rules.

For example, tripwires generally perform better than, for example, motion detection. This is because they are placed in controlled environments, typically along a fence, where there are very few alternative explanations as to why an object would cross the fence without posing a threat.

Object left behind, on the other hand, is a much more challenging algorithm. This is because an object might not be left behind at all, and most objects left behind pose no real threat. For example, garbage thrown onto train tracks is not a bomb, but would still be seen as an object left behind. You have to ask yourself the right questions, such as what is the condition you are trying to detect, to pinpoint the right analytics system, the right camera setup, and the right parameters.

Ideally, analytics should be able to differentiate between true-threat events and posing-as-threat events. For example, in a retail scenario, it would be ideal if the same application could not only detect that a customer put something in their bag, but also that they went through the self-checkout area without scanning and paying. It is not until then that they can be accused of shoplifting.

On a technical level, four important factors must be present for an intelligent video application to work accurately:

1. The right input quality (i.e., video image quality)
2. Efficient intelligent video algorithms
3. Ample computer processing power
4. Configuration and fine-tuning

### 16.7.1 Video image quality

One of the steps to make an intelligent video application work well is getting the right video from the cameras, that is, video suitable for processing in that particular application. Some important considerations include the following:

- *Lighting and the ability to see in darkness:* Any type of video analytics application is limited to what the camera actually can see. Is there sufficient light and does it come from the right direction? What happens if the camera is blinded by headlights? Can the camera counteract bad lighting conditions with WDR or low-light capabilities? If detection is needed in complete darkness, a thermal camera is likely the best option.
- *Frame rates and resolutions:* Contrary to popular belief, most analytics do not need high frame rates and high resolutions. In fact, 5–10 fps in 4 CIF (roughly 0.4 megapixel) resolution is sufficient for most general intelligence applications. Some specialized applications, such as LPR and facial recognition, may need even lower frame rates, although they work better at higher resolutions.
- *Camera position:* Positioning cameras is critical to getting accurate results from analytics. For example, most people-counting applications work best with cameras positioned overhead because this allows the algorithms to efficiently separate individuals. In such use cases, because overhead cameras are not optimal for other types of detection, it is best to use dedicated cameras. Applications such as LPR and facial recognition work best with frontal views, that is, with the camera looking straight at the license plate or face.
- *Type of video:* Because compression always results in loss of data, the best results are achieved when processing uncompressed video, as is often the case when running video analytics on the edge, inside the network camera. Some intelligent video applications that require object classification need color video. Others that simply count or track can use black-and-white video. A very interesting possibility is running analytics based on data from thermal cameras. Thermal cameras are perfect when you need to tell living objects from dead objects or detect activity in completely dark areas or when you have to cover long distances (thermal cameras have longer detection ranges than conventional cameras).

### 16.7.2 Efficient intelligent video algorithms

Intelligent video applications are built on complex mathematical algorithms that process video and still images, each consisting of a myriad of details. The quality of an intelligent video application depends on how accurately the algorithm performs these calculations and how robustly it deals with variations in the input (i.e., the video stream). The only sure way of assessing the quality of an intelligent video algorithm is to field test the application under realistic conditions and see how fast it is and how many correct responses and false alarms it generates.

In general, 90% accuracy is achievable for a modern intelligent video system. Reaching 95%, however, is very complicated, and 99% or beyond is extremely difficult in a real-world situation. From a user's point of view, the demand for accuracy depends on the following examples: How critical is the analytics application for the safety and security of people and property? How many errors are acceptable? How many false alarms can the system be allowed to generate before it is unusable? How many true positives (situations that should generate alarms) can the system be allowed to miss? The cost of an intelligent video system must also be weighed against other alternatives, such as employing more security personnel.

### 16.7.3 Computer processing power

Because analytics applications are mathematically complex and therefore computer-power intensive, performance depends on the processors and available memory. Some intelligent video applications are optimized to run on small, embedded systems and perform well in a distributed system, whereas others require a very powerful centralized server to be reliable. In either case, the more processing power available to an intelligent video application, the better and faster it will be.

### 16.7.4 Configuring and fine-tuning the system

Intelligent video applications and algorithms are designed to handle a large variety of situations. Every installation requires configuration and fine-tuning to match each particular scenario. No system is perfect, and configuring a system is a balance between not missing essential situations and reducing false triggers. Optimizing a system can take anywhere from a day up to several weeks.

As a rule of thumb,

- The system will never be 100% accurate.
- The more parameters that can be adjusted in an application, the longer it takes to optimize.
- Monitor and adjust the configuration over at least one 24-hour period, as changes in lighting will impact results.
- Using thermal cameras can increase accuracy dramatically in some applications, such as crossline detection.
- Combining video analytics with other subsystems such as access control can increase accuracy.



## CHAPTER 17

# System design considerations

One of the main benefits of a network video system is flexibility and scalability: the freedom to mix and match the most appropriate components from different vendors and the power to optimize or expand the system to any size. As with everything else, having freedom and power also demands knowledge. To build a truly flexible and scalable system, you need to know what the different components are, how they work, and how they interact. It is essential that you can select the right camera, install and protect it properly, configure it to match the scene complexities, and get it to stream live or record video at the right time, in the right format, and with the right quality. At the same time, the appropriate network and storage solutions depend greatly on the selected cameras, the camera settings (such as resolution, compression, and frame rate), and the number of cameras. The budget might demand compromises. For example, as we already discussed, complex scenes demand more bandwidth, so you may have to choose between spending more on storage, reducing the frame rate, lowering the resolution, or increasing the compression.

This chapter discusses the most important aspects of designing and installing a network video system: how to select, install, and protect a network camera and how to calculate the storage and network bandwidth. Many design tools are available, some of which are based on formats such as AutoCAD® or Revit®, which will also be discussed. There are also legal aspects to consider, some of which are mentioned at the end of the chapter.

### 17.1 SELECTING A NETWORK CAMERA

---

In growing markets such as the video surveillance and physical security market, new vendors are always appearing, bringing new products that potentially also have new types of abilities. Today, there are hundreds of different network camera vendors in the marketplace. Because network cameras include much more functionality than analog cameras, choosing the right camera becomes not only more important but also more difficult. When choosing the technology to use, you need to consider that many physical security systems remain in operation for 5–10 years, during which time they need to be maintained and serviced. This section outlines what to keep in mind when selecting a network camera. This includes the type of camera, image quality, resolution, compression, networking, and other functionalities, as well as the vendor.

### 17.1.1 Types of camera

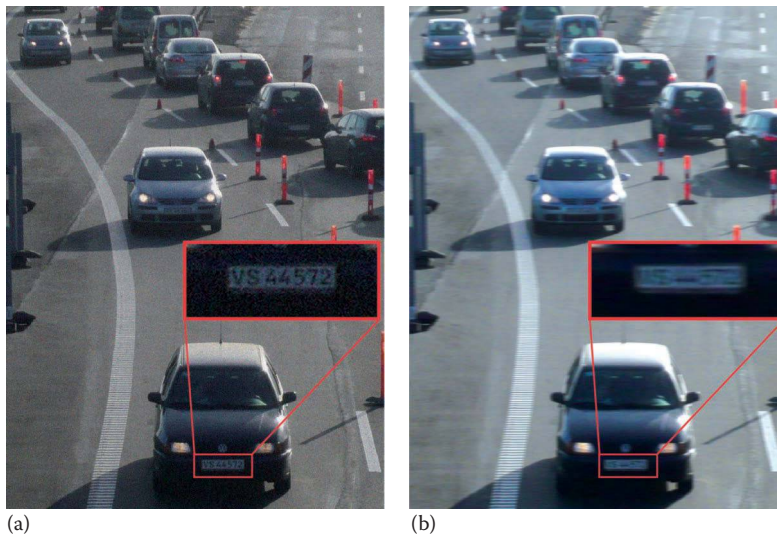
To determine which types of network cameras are suitable and how many cameras are needed to cover an area, the scene, environment, and purpose must be determined first. Considerations include the following:

- *Indoor or outdoor camera:* If placing the camera outdoors, install it in an appropriate protective housing or use an outdoor-ready camera. Look for the IP rating (IP66 or better) or National Electrical Manufacturers Association (NEMA) rating (4X or better). The camera should also have auto-iris functionality.
- *Pan, tilt, and zoom (PTZ) or fixed camera:* PTZ cameras with a large optical zoom factor can give high-detail images and survey a big area. Keep in mind that to make full use of the capabilities of a PTZ camera, an operator needs to control the movements, or an automatic tour must be set up. For surveillance recordings without live monitoring, fixed network cameras are normally more cost effective.
- *Light-sensitivity and lighting requirements:* Consider adding external white lights or specialized lighting such as IR (infrared) lamps. Day-and-night functionality means you can get images in conditions that would otherwise be too dark. The light-sensitivity levels of a camera are important and should be evaluated. Do not go by the measurements on a datasheet as vendors measure in different ways. For more information on light-sensitivity measurements, see Section 4.1.2.
- *Complete darkness and perimeter protection:* Thermal cameras can detect movement even in complete darkness and other difficult conditions. Generally, they can also detect movement at greater distances than conventional cameras.
- *Tamper- or vandal-proof and other special housing requirements:* Proper protection against water, dust, temperature, and vandalism is essential. For more information, see Section 17.3.
- *Overt or covert surveillance:* This will help in selecting cameras that offer a nondiscreet or discreet installation.
- *Area of coverage:* For a given location, determine areas of interest, how much of these areas should be covered, and whether the areas are located relatively close to each other or spread far apart. For example, if there are two relatively small areas of interest that are close to each other, a high-resolution camera with a wide-angle lens can be used instead of two cameras with lower resolution.
- *Overview or high-detail images:* Determine also the field of view or the kind of image that will be captured: overview (viewing a scene in general or looking at the general movements of people) or high detail for identification of persons or objects (e.g., face or license plate recognition, point-of-sale [POS] monitoring).
- *Analytics:* Camera type and placement affect the success of most video analytics software and the other way around. Also, consider whether the system will work best with edge-based or server-based analytics. For more information, see Chapters 15 and 16.

### 17.1.2 Image quality

Although image quality is one of the most important aspects of any camera, it is difficult to choose the right camera based on a datasheet. The reality is that many aspects of image quality cannot be quantified or measured. To illustrate the challenge, consider the two images in Figure 17.1. The images were taken from two different cameras with same resolution and similar specifications, under the same conditions and illumination. The cameras cost about the same and come from brand-name vendors. The conclusion is that the best way to determine image quality is to install different cameras and look at the video. Keep in mind that although a camera may provide good still images, the images may not be as good when a lot of motion is introduced into the scene.

Many factors affect image quality. For example, white balance and a camera's ability to adapt to different lighting conditions from fluorescent, high-pressure sodium, to LED light is important to



**Figure 17.1** Two similar cameras recording the same scene can give images with very different quality. One image has a lot of noise, but it provides enough detail to make positive identification (a). The other image is more pleasant to look at, but the distinguishing details are lost due to motion blur (b).

ensure color fidelity. Low-light, backlight, dynamic light, and other extreme lighting conditions present challenges that the camera needs to be able to handle. Typically, a high-resolution camera is less light sensitive than a lower-resolution camera. In other words, you may need to consider sacrificing resolution for better low-light performance, or use a camera with a sensor and processing algorithms that are especially designed to meet these challenges. For more information about how light sensitivity, image processing, scanning techniques (such as progressive scan), sensor size, and other factors affect image quality, see Chapter 4.

### 17.1.3 Resolution

Best practices have emerged regarding the number of pixels required for certain video surveillance operations. For an overview image, 70–100 pixels are generally enough to represent 1 m (20–30 pixels/ft) of a scene. For operations that require detailed images, such as face recognition, the demands can increase to as many as 500 pixels/m (150 pixels/ft). This means that you need to be able to generate positive identification of people passing through an area 2 m wide  $\times$  2 m high (7  $\times$  7 ft), the camera needs to provide a resolution of at least 1 megapixel (1000  $\times$  1000 pixels).

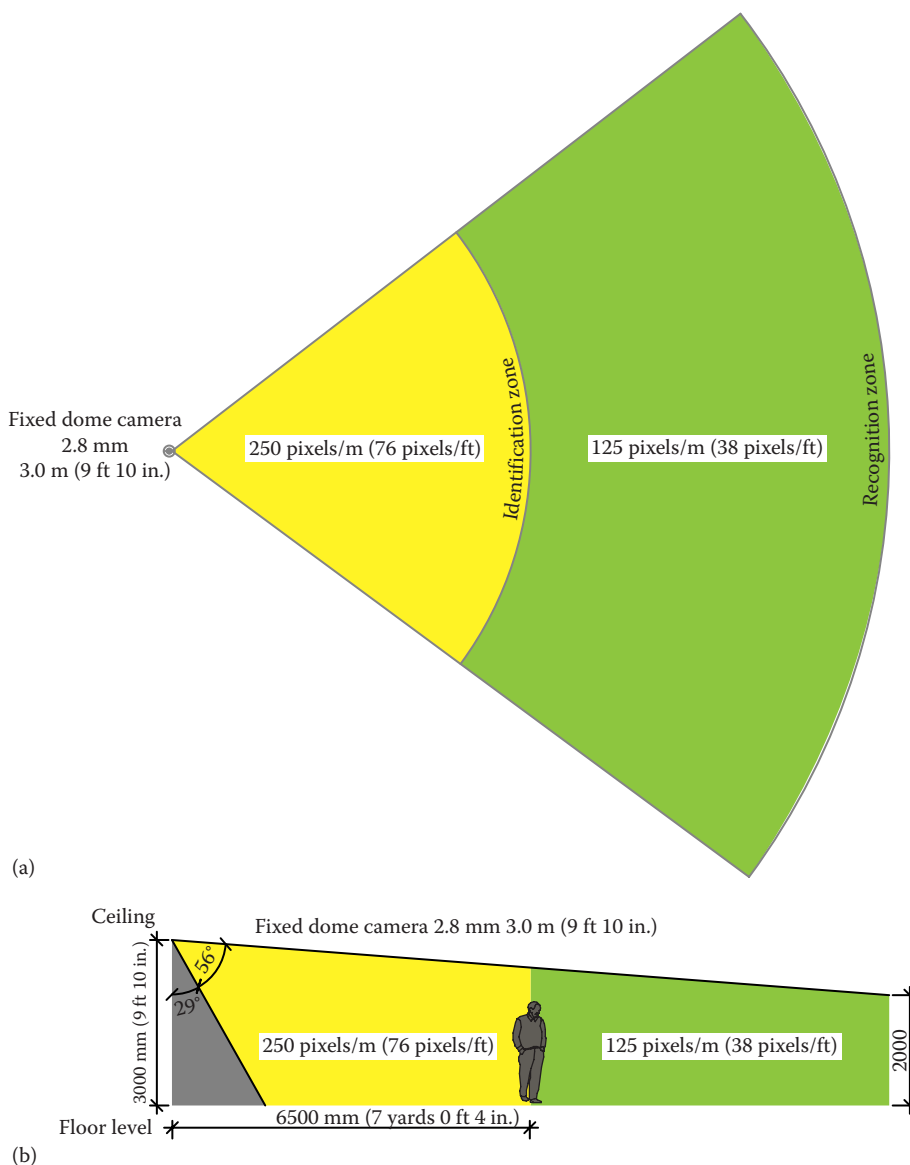
Figure 17.2 illustrates the field of view of a fixed dome camera with a 2.8 mm lens. The camera is mounted 3 m (9 ft 10 in.) above the ground. The green area gives a resolution of 125 pixels/m (or 38 pixels/ft), which is the limit for recognition, up to 250 pixels/m (or 76 pixels/ft), which is the limit for identification.

The maximum distance for identification and recognition using different camera models with different lenses is shown in Figure 17.3. Again, there are tools that help calculate resolution and field of view based on camera model and position; see Section 17.6.

#### 17.1.3.1 Determining the resolution needed

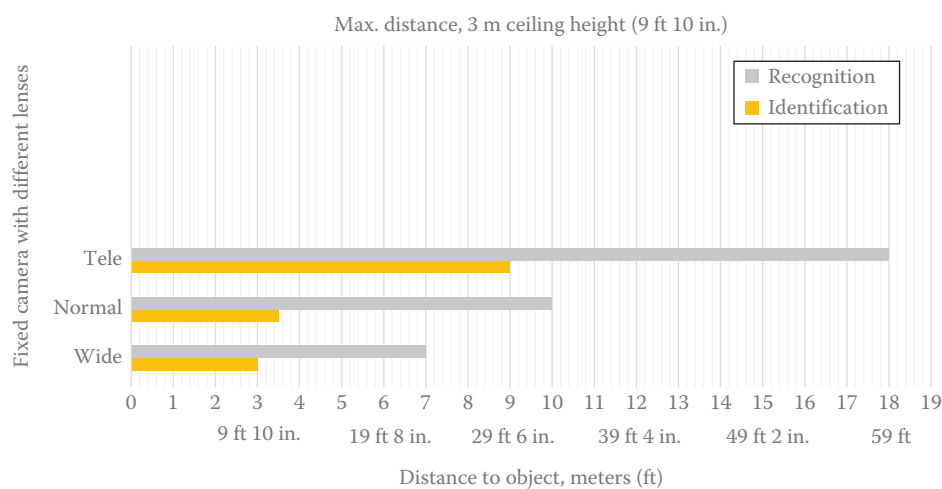
The required resolution for video surveillance images depends on the size of and the distance to the objects under surveillance. To illustrate this, consider a surveillance of an airport entrance where people need to be identified any time there is an incident.

Say the entrance area is 20 m (65 ft) wide. The estimated vertical field of view needed to identify a person is 2 m (6 ft). But because network video offers a multitude of resolutions and formats,



**Figure 17.2** A person standing 6.5 m (21 ft 4 in.) away can be identified when the camera is mounted at a height of 3 m (9 ft 10 in.). The horizontal coverage of the camera is shown in (a) and the vertical coverage in (b).

it is better to define resolution requirements based on pixel density in the horizontal dimension. The most typical object that needs to be recognized is a human face. The variances in face width are less than those of body length or width, so there is a smaller margin of error. Depending on the conditions, you need 40–80 pixels across the face for positive identification of a person. Some manufacturers and organizations, such as Swedish National Laboratory of Forensic Science (SKL), recommend 80 pixels across a face for identification, while others, such as the international standard published by CELENEC, suggest that 40–100 pixels/ft is sufficient. Counting pixels across the face is a convenient measuring method because all adult faces are about the same width, which is about 0.16 m or a little more than  $\frac{1}{2}$  ft. Therefore, in the following examples, we focus on the horizontal resolution. The vertical resolutions that correspond to the 2-m vertical field of view are included for reference.



**Figure 17.3** An example of the maximum distance for identification and recognition with different camera models and lenses. For example, the camera with a telelens enables identification at up to 9 m (29 ft 6 in., or almost 10 yards) and recognition at up to 18 m (59 ft, or approximately 19 yards).



**Figure 17.4** The effect of increasing resolution. These images show what 80 (a), 40 (b), and 20 (c) pixels across the face can look like.

Figure 17.4 shows images of a face using three different resolutions, measured in number of pixels across the face. You can clearly see the impact of higher resolutions (see also Tables 17.1 and 17.2).

If you know the number of pixel needed across the face (face resolution) and the width of the surveillance area (scene width), you can calculate the total number of pixels needed to cover the width of the scene (horizontal scene resolution). Divide the scene width by the face width (0.16 m or ½ ft), and then multiply with the face resolution:

$$\text{Horizontal scene resolution} = \frac{\text{Scene width}}{\text{Face width}} \times \text{Face resolution}$$

In Table 17.1, we use the measurements from the airport entrance scenario mentioned earlier.

**Table 17.1** Required number of pixels per scene, given that the width of a typical face is 0.16 m (½ ft)

Alternative	Face resolution, W (pixels)	Scene, W × H	Scene resolution, W × H (pixels)
A	20	20 m × 2 m	2500 × 250
B	40	20 m × 2 m	5000 × 500
C	80	65 ft × 6 ft	10,400 × 960

**Table 17.2** Number of cameras needed to cover the scene

Alternative	Scene resolution, W × H (pixels)	720p (1 megapixels) 1280 × 720	1080p (2 megapixels) 1920 × 1080	4K (8 MP) 3840 × 2160
A	2500 × 250	2	2	1
B	5000 × 500	4	3	2
C	10,000 × 1000	8	6	3

Calculation in meters, alternative B:

$$5000 \text{ pixels} = \frac{20 \text{ m}}{0.16 \text{ m}} \times 40 \text{ pixels}$$

Calculation in feet, alternative C:

$$10,400 \text{ pixels} = \frac{65 \text{ ft}}{1/2 \text{ ft}} \times 80 \text{ pixels}$$

With the resolutions determined, you can calculate the number of required cameras:

$$\text{Number of cameras} = \frac{\text{Horizontal scene resolution}}{\text{Camera resolution}}$$

Table 17.2 shows some possible camera combinations and the benefit of megapixel cameras.

Although somewhat hypothetical, the examples in Table 17.2 show that there is a lot to gain by calculating how many cameras you need. If the plan is to get high-resolution coverage of this scene with alternative A, you need eight 720p cameras. This drops to three cameras if you use 4K-cameras instead—a number that is easier to handle. Likewise, if the demands on resolution are moderate, there is little point in going for a 4K-camera if a 720p-camera will do.

Some cameras have a built-in pixel counter, which makes it easier to verify that the camera’s resolution is high enough for the scene and to make a positive identification (see Figure 17.5). It is not necessary to use a model subject when you calculate the pixels; a ruler or other object with a known width



**Figure 17.5** Examples of a built-in pixel counter. To change the size of the selection frame, you either drag the corner or set the required width and height in pixels. Are there enough pixels to identify a person when they stand in this part of the scene (a)? Are there enough pixels across the counter so that when a person stands there definitely can be identified (b)?



(such as a sheet of copy paper) works too. If the object is too small to fit in the pixel counter frame and you cannot accept a smaller field of view, you need to get a camera with a higher vertical resolution.

### 17.1.4 Compression

The most common types of video compression are Motion JPEG and H.264. As discussed in Chapter 6, we can expect H.265 to become established in the coming years, at least in the consumer market and possibly in the video surveillance market as well. Until then, much can be gained by using improved H.264 for surveillance video. Each format has its advantages and disadvantages. When network cameras offer support for more than one type of video compression standard, they give users greater flexibility in optimizing viewing and recording needs. For example, some systems may want to use Motion JPEG for live viewing and H.264 for recording.

Designing the network and storage system is highly dependent on the selected compression standard. The following are some basic recommendations:

- *Motion JPEG*: This is suitable for smaller systems with limited retention requirements, systems that only need a few frames per second, and systems with remote cameras. It consumes more bandwidth and storage. It is often used in low-frame rate scenarios to capture high-quality images, not video, such as LPR applications where the vehicle is moving at a high speed. Even at low frame rates, MJPEG can capture images of such scenes with more detail than H.264.
- *H.264*: H.264 is the dominant video compression standard in video surveillance today. Compared to MJPEG, H.265 reduces network bandwidth and storage. It captures video at a higher rate with more efficiency. The Main and High profiles have made H.264 more efficient and have improved image quality too.
- *Improved H.264*: The H.264 standard only refers to the decoder, whereas there are opportunities to improve the encoding. New technologies can filter out low-interest areas and compresses them more, while recording the details of interest and motion at higher quality. This switching between a maximum GOP value and a lower value drastically reduces the bitrate and the requirements on storage and bandwidth.
- *H.265*: Network video products with H.265 compression have entered the market, but they are still scarce and not as efficient as the improved H.264 mentioned earlier. H.265 provides high resolution at lower bitrates, lowers the minimum bandwidth requirements, and reduces storage needs. However, it requires high-performance network cameras with sufficient computing power and efficient decoders on the video monitoring side of the system.

For more information on compression, see Chapter 6.

### 17.1.5 Networking functionality

In the same way that high-quality images are essential, networking functionality is also important. Besides Ethernet connectivity, which is obvious, a professional network camera should also support the following capabilities:

- *Power over Ethernet (PoE)* means that the camera can receive power through the same cable as for data. PoE eliminates the need for power cable runs. This means you can save hundreds of dollars per camera in installation costs alone. Make sure the camera complies with the PoE IEEE 802.3af/802.3at Type 1 (up to 15 W) or 802.3at Type 2 (up to 30 W) standard. This will give you the freedom to select from a wide variety of network switches. Outdoor PTZ cameras may require more power (up to 60 W or even more in extreme environments), which can be provided by a midspan. For more information about PoE, see Section 9.6.
- *Dynamic Host Configuration Protocol (DHCP)* is used by many organizations to manage IP addresses. A DHCP-enabled switch or router automatically gives each connected device an IP address. Some advantages of that include its quicker and easier ability to deploy or switch cameras, stability through periodic renewal, reservations, and failover. With less advanced or

poorly managed DHCP servers, the disadvantages include the risk of temporarily losing IP addresses if the server goes down or resets, that IP addresses change over time, or that someone accidentally connects the server to a router that is open to the internet. Some prefer the predictability of static IP address and the ability to match the IP address with the cameras' ID numbers.

- *HTTPS encryption* for secure communication.
- *SNMP* helps IT administrators monitor conditions of the network and determine which connected devices might need immediate attention.
- *IP address filtering* enables only defined IP addresses to have access to the camera.
- *IPv6* is the most recent version of the IP. IPv6 uses 128-bit addresses that are divided into eight groups of four hexadecimal digits. Its development was stimulated by the limitations of IPv4, which uses 32-bit addresses.
- *Wireless technology*: Wireless is a good option if running a cable between a LAN and a network camera is impractical, difficult, or expensive. Wireless access also can be provided for a standard network camera, but it must first be connected to a wireless device point. Wireless technology can be useful, for example, in historic buildings where the installation of cables would damage the interior, or in facilities where there is a need to move cameras to new locations on a regular basis, such as in a supermarket or in outdoor installations. Wireless technology can also be used to bridge sites without expensive ground cabling. For more information about wireless networks, see Chapter 10.
- *802.1X* enhances the security of wireless local area networks through a port-based authentication framework.

Always weigh in the opinion of the IT department. They should be able to determine if the camera provides adequate networking functionality and security. For more information on networking technologies, see Chapter 11.

### 17.1.6 Other functionalities

Network cameras have many other functionalities apart from just providing a video stream. When selecting a camera, it is important to evaluate these capabilities too. Some examples of additional functionalities are listed here:

- *Audio for communication and detection*: Users can listen in on an area and communicate instructions, orders, or requests to visitors or intruders. When microphones detect sounds above a certain level, they can trigger alarms or cameras to start recording. Consider whether one-way or two-way audio is required. Microphones and speakers can be built-in or external. Consider the legal aspects of monitoring and recording audio in your region. Some regions forbid the use of audio for surveillance purposes except in police interview rooms and other government-controlled facilities. For more information about audio, see Chapter 7.
- *Built-in analytics*: Many network cameras offer built-in intelligence such as video motion detection, tampering detection, and people counting. Built-in intelligence makes the system more scalable and helps reduce bandwidth and storage requirements because the camera is able to decide when to send and process video. Analytics require a lot of processing power, and if the processes are performed on the PC server rather than in the camera, the PC can quickly become overloaded. For more information about analytics, see Chapters 15 and 16.
- *Input/output (I/O) connectors*: Connecting external input devices to a camera (such as a door contact, infrared motion detector, glass-break sensor, or shock sensor) enables the camera to react to an external event by, for example, sending and recording video. In many cases, as in scenarios where the goal is to capture the identity of a person at an entrance, there is no need for the camera to continually send video. Through the input port, the camera knows when the door opens and only captures and sends video of that particular event. Outputs enable the camera or a remote operator to control external devices, for example, alarm devices, door locks, or lights.
- *Alarm management functions*: Advanced network cameras can perform alarm management tasks, such as processing and linking input, output, and other events. For example, if the level

of audio in a room passes the threshold, the camera can send an output signal that turns on the lights and send video to the video management software. Pre- and postalarm buffers in a network camera can record video before and after an alarm occurs. After detecting an alarm or event, a network camera can send notifications via email, TCP, and HTTP and upload images through FTP, HTTP, and email or directly to network attached storage (NAS) devices and SD cards.

- *Other physical security devices:* In an enterprise environment, it is favorable to take a holistic approach to the system requirements. When you evaluate a video surveillance system, also look at other systems, such as access control, intercom, audio, and intrusion detection, to determine if there is a way to build out an integrated system that covers all physical security needs.

### 17.1.7 Vendor

Choosing a video surveillance vendor and partner can be confusing when there are so many to choose from, all with their own range of products and solutions. Narrow down the selection by isolating one or a couple of vendors. Here are a few tips that may help when selecting a network camera vendor:

- *Wide product portfolio:* Go with those who maintain a full product line, including fixed cameras, fixed dome cameras, and PTZ cameras. This way, one or two companies can satisfy current and future needs for system expansion and functionality upgrades such as megapixel, wireless, or audio. If analog video products must be integrated into a network video system, make sure that the chosen company's product portfolio also includes video encoders and decoders.
- *Multiapplication support and ease of integration:* Make sure you select network cameras that have open application programming interfaces (APIs) and can integrate with several video management software applications. Some network camera vendors have hundreds of such alliances. Open multivendor video management systems (VMSs) give users the most flexibility.
- *Tools for managing large deployments:* Like all intelligent network devices, network cameras have an IP address and built-in firmware. Upgrading the firmware is usually easy and many vendors provide upgrades free of charge. When making a purchase decision, consider the cost of setting IP addresses and future upgrades of all the cameras in the system. The network camera vendor should have tools to manage these processes, and their estimates for cost and downtime should be clear and measurable up front. The vendor should also have software that can locate all network video devices automatically and monitor their status.
- *Tools and support for system design:* Enterprise-level systems have tremendous capabilities and flexibilities. This also means that to get the most out of the system, it needs to be properly designed. Does the vendor have the support and tools to help with designing and documenting the system?
- *Warranty:* Video surveillance systems are a substantial investment for most organizations and have a life expectancy of several years. Make sure that the vendor has a reasonable warranty on the selected cameras, and ask if they offer extended warranty services.
- *Networking knowledge:* In the analog video surveillance world, there was little need to evaluate the IT knowledge of video surveillance vendors. In the world of IP-based security and surveillance systems, the case is very different. Not only is the networking functionality important, but even more so the technologies available to provide adequate cybersecurity.
- *Long-term partner:* Select a vendor that has the potential of being a long-term partner. Remember that your system needs to be operational, maintainable, and perhaps expandable for 5–10 years. Does the company have a large base of installed cameras? Is their focus on network camera technology? Can they offer local representation and support? Is the company a global player? Because needs change and grow, it is important to choose a vendor where there is innovation and long-term plans for support, upgrades, and product paths. Look at the prospects of future growth and the need for added features and functionality.

After making a decision as to the desired camera, it is a good idea to purchase a single unit and test its quality and real-world performance before ordering large quantities.

## 17.2 INSTALLING A NETWORK CAMERA

How a network camera is installed is just as important as the process of purchasing it. Discussed in the following are some recommendations on how to best achieve high-quality video surveillance based on camera positioning and environmental considerations.

### 17.2.1 Surveillance objective

To best position a camera, you need to know what kind of image you need. For example, to track people or objects moving to and from many positions in several directions, you probably want an overview image as it gives the best chance of spotting such events. After finding and purchasing a suitable overview camera, you must install it in a position that achieves the purpose.

To identify a person or object, the camera must be positioned or focused so it can capture the level of detail needed for identification purposes. As seen in Figure 17.6, the larger the angle to the object, the more difficult facial features are to recognize. An angle of 10°–15° gives the best view for facial identification. On the other hand, placing a camera higher up puts it out of reach for vandals, but again, the bird's-eye perspective makes identification of faces or details, such as license plates, more difficult.

There are tools that can help find the best position for a camera. Local police authorities may provide video surveillance guidelines. A spinning Rotakin can be used to test how well a camera displays moving objects (see Figure 17.7).

Some vendors also provide design tools and plug-ins for various diagram and 3D software. These tools can help with placement of cameras, calculating view angles and coverage, and finding blind spots and items that block the view (see Section 17.6).

### 17.2.2 Use plenty of light or add light if needed

The most common reason for poor quality images is lack of light. An easy, cost-effective way to improve the lighting conditions and get better images is to add more light. This is a solution that works in many indoor and outdoor situations. The positioning of external lights is as important as



**Figure 17.6** The greater the angle to the object, the more difficult facial features are to recognize.



**Figure 17.7** Spinning Rotakin.

the positioning of the camera. You want to avoid reflections, shadows, or blinding the camera or people moving in the area.

In basic terms, there are three situations in which adding light would help:

1. The scene is too dark for the camera to produce useable images.
2. The lighting of the scene is not good enough for the camera to produce high-quality or bright images without adding either motion blur or noise. Remember, noisy images consume more bandwidth and storage.
3. The lighting conditions of the scene are challenging. For example, the scene includes both bright and shadowed areas, or the scene suffers from backlighting. A consistently even-lit scene is always easier to deal with, no matter if it is dim or very bright.

How to position a white-light illuminator relative to the camera's position matters, especially when areas in the scene have different reflective ability depending on factors such as the weather. For example, worn, light-gray asphalt becomes completely dark when wet.

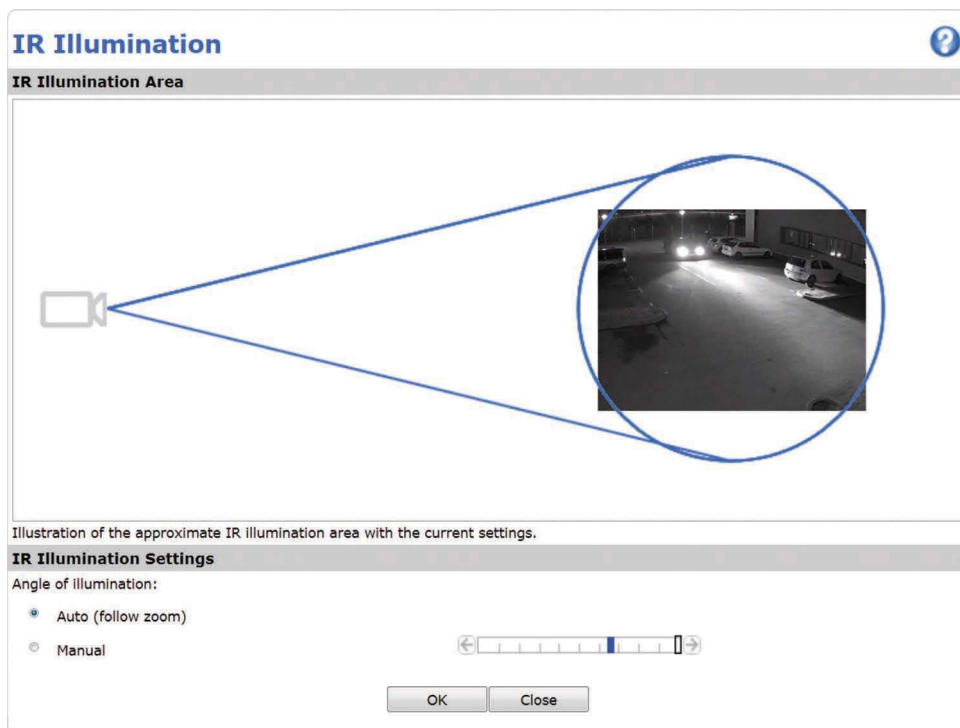
Smooth and textured surfaces reflect light differently. Uneven surfaces bounce light in all directions because of their inherent irregularities. When the light hits the object squarely, the reflection tends to be stronger. So typically, the camera's position should be beside the illuminator and directed straight at the target. This also helps the camera deal with shadows in the scene.

With smooth surfaces, the angle of reflected light is equal to the incoming light. In these cases, the camera and the illuminator should be placed so they have the same angle. However, smooth surfaces are uncommon in video surveillance.

#### 17.2.2.1 Use IR light when white light is impracticable

For discreet or covert security or in areas where the presence of artificial light is unwanted, choose an IR-sensitive day-and-night camera. An IR illuminator, which provides near-infrared light, can be used in conjunction with an IR-sensitive camera or a day-and-night camera to enhance the camera's ability to produce high-quality video in low-light or nighttime conditions. To maintain good image quality and prevent focus shifting at night when using IR illuminators, make sure that the camera has an IR-corrected lens. For more information about IR illuminators, see Chapter 3.

There is also the option of using cameras with built-in IR LEDs. Previously, built-in LEDs had a very short lifespan, and they were known to generate heat that would increase the noise in the image. Recent technology advancements have resulted in new mechanical designs, in which heat



**Figure 17.8** In network cameras with state-of-the-art built-in IR LEDs, the angle of illumination can follow the camera's field of view automatically.

can escape more easily; a longer lifespan mirrors that of the camera, with gaskets that prevent IR light from leaking in and polluting the images; and autofocus follows the camera's focus and field of view (see Figure 17.8). With built-in LEDs, it is possible to configure the camera to turn the lights on automatically at a specific time, when the amount of ambient light reaches a specific level, or when triggered by an event. Using built-in IR LEDs is straightforward because of the seamless design and ease of installation.

### 17.2.3 Avoid direct sunlight and glare

Always avoid direct sunlight into the camera. Direct sunlight blinds the camera and can reduce the performance of the image sensor. If possible, position the camera with the sun behind the camera. Some cameras have sunshields that help reduce the impact of direct sunlight and make placement easier.

Sunlight is not the only issue. A camera can usually deal with being pointed toward a self-advertising lightboxes during the day, but at night, the ambient light level is lower and the contrast between high and low light levels is greater. Therefore, objects such as car headlights or illuminated signs appear brighter than they do in daylight (see Figure 17.9). Such glare is problematic for cameras and human eyes alike. To help the camera get better video, increase the ambient light level in the scene by using additional lighting.

### 17.2.4 Avoid backlight

Backlight creates various problems for all conventional cameras. Here are some of the most common problems and suggestions on what to do about them.

- *Avoid the brightest areas:* Pointing or focusing the camera at the brightest areas in the scene causes problems with over- and underexposure. The bright areas become overexposed (bright white), while other objects appear too dark. This problem typically occurs when attempting to





(a)



(b)

**Figure 17.9** Two images of the same scene, taken in different lighting conditions. During the day, the colors are vibrant (a). At night, the contrast between light and dark makes it difficult to see the area surrounding the lightbox (b). Adding some ambient light would improve the image quality.

capture an object in front of a window. To solve this problem, reposition the camera or block the light with curtains or blinds. If neither is possible, add frontal lighting.

- *Reduce the dynamic range:* In outdoor environments, viewing too much of the sky results in too high a dynamic range. The camera will self-adjust to achieve a proper light level for the sky. Consequently, the object or landscape of interest will appear too dark (see Figure 17.10). One way to solve this problem is to mount the camera high above the ground, using a pole if needed.
- *Adjust the camera settings:* It may be necessary at times to adjust settings for brightness, sharpness, white balance, and wide dynamic range (WDR) for different environments (indoor, outdoor, and fluorescent) to get an optimal image. Some areas have multiple light sources, for example, fluorescent light fixtures together with daylight. Make sure the camera can deal with this.

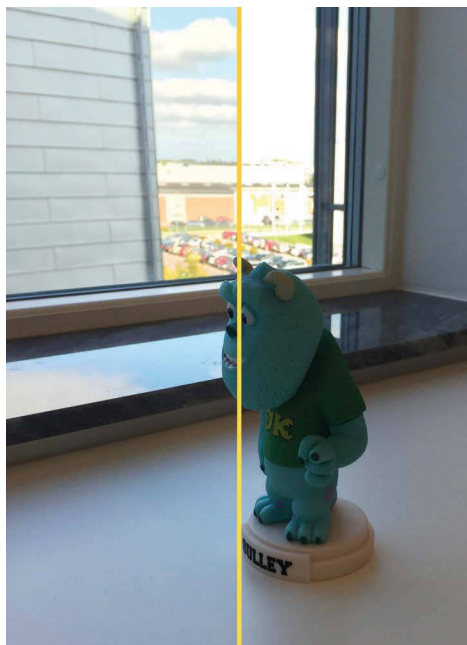


(a)



(b)

**Figure 17.10** To avoid very bright areas in an image, such as the backlight shown here (a), change the camera position (b).



**Figure 17.11** Advanced cameras include a feature that compensates for backlight, which in this example is most apparent in the sky part of the image.

Cameras with support for a high (or wide) dynamic range are better at handling backlit scenarios (Figures 17.11 and 17.12). As discussed in Chapter 4, there are several different types of WDR technologies.

- *Change the exposure time:* Use short exposure times when rapid movement occurs or when a high frame rate is required. A long exposure time makes the images look nicer, but probably lowers the total frame rate and results in increased motion blur. In network cameras with automatic exposure, the frame rate increases or decreases with the amount of available light. It is only as the light level decreases that artificial light or prioritized frame rate or image quality becomes an issue. As newer cameras have better low-light abilities thanks to improvements in sensor technology and image processing, the trade-offs between motion and noise are less than they used to be.

### 17.2.5 Lens selection

The fastest way to calculate the required lens and field of view is to use an online lens calculator, preferably one that also considers pixel density, as available from most camera and lens manufacturers. Alternatively, use a rotating lens calculator (Figures 4.16 and 4.17). For more information on lenses, see Section 4.2.

For outdoor situations, it is best to use an auto-iris lens because it automatically adjusts the amount of light that reaches the image sensor. This optimizes the image quality and protects the image sensor from being damaged by strong sunlight. Even better than an auto-iris lens is a P-Iris lens, especially when using high-resolution cameras. P-Iris lenses are less prone than standard auto-iris lenses to producing diffraction in their images. Diffractions are especially an issue when the light is strong. And the higher the resolution of the sensor, the more prominent the diffractions. For more information about the differences between auto-iris and P-Iris lenses, see Section 4.2.6.

Long-distance coverage usually requires longer lenses than those included with the camera, so this is something to consider when installing a camera in an outdoor housing—will the lens fit in the housing?

**White Balance**

White balance: Automatic Edit...

White balance window: Automatic Edit...

**Wide Dynamic Range**

☒ Enable Dynamic Contrast

Low High

**Exposure Settings**

Exposure control: Automatic

Max exposure time: 1/35500 s

Enable Backlight compensation ☒

Exposure zones: Auto Defined [Auto] Edit...

**Shutter & Gain**

Shutter: Auto

Gain: Auto

**Normal Light**

Priority: Low noise Low motion blur

Max gain: 21 dB

Max fast shutter: 1/1000 s

**Low Light**

Priority: Low noise Low motion blur

Max gain: 42 dB

Max shutter: 1/30 s

☒ Enable automatic iris adjustment

Iris adjustment: F 1.2 71 [0..100]

**Day/Night**

IR cut filter: Automatic

Day/Night shift level: Sun

☒ Enable focus adjustment for IR illumination

**Figure 17.12** An example of a user interface with options for how the camera should handle light.

## 17.3 PROTECTING A NETWORK CAMERA

Surveillance cameras are often placed in very demanding environments. In outdoor installations, protection against varying weather conditions is necessary. In industrial settings, cameras may require protection from hazards such as dust, acids, or corrosive substances. In vehicles such as buses and trains, cameras must withstand high humidity, dust, and vibrations. Cameras may also require protection from vandalism and tampering.

Manufacturers of cameras and camera accessories use various methods to meet environmental challenges. Solutions include placing cameras in protective housings, designing special-purpose cameras for each type of environmental challenge, and using intelligent algorithms that can detect and alert users to a change in a camera's operating conditions.

The level of protection provided by enclosures, whether it is a one-fits-many or an integrated solution, is often indicated by IP, NEMA, and IK ratings. All electronic devices, cameras, and encoders

must also fulfill the emission, immunity, and electronic safety requirements of the region and the environment in which they are used.

The following sections discuss such topics as coverings, positioning of fixed cameras in enclosures, environmental protection, vandal and tampering protection, types of mountings, and protection ratings.

### 17.3.1 Camera enclosures in general

A camera's operating conditions depend on its materials and components. When the demands of the environment are beyond a camera's operating conditions, the camera needs to be protected by a housing (also called an enclosure). Some cameras are designed for outdoor and other demanding conditions and have a built-in enclosure. However, to use a camera designed for indoor and less demanding environments, you need a separate enclosure in which to place the camera.

Camera enclosures come in different sizes and qualities; some are made of metal, others of plastic. Most cameras today come with the protective enclosure as an integrated part of the camera. When selecting the appropriate enclosure for a specific camera, several things must be considered (Figure 17.13):

- *Mount*: What kind of mounting bracket do you need (wall, pole, corner, parapet)?
- *Cable runs*: How can the cables be run? How are you going to manage them? How much cable do you need? What kind of shielding is required? What quality of cable do you need for your data? What kind of conduits do you need?
- *Operating temperature and environment*: Do you need heaters, fans, sunshields, or wipers. Do you need or dust-proof or water-proof materials and seals?
- *Power supply*: How much power does the camera need? Does the housing need power? What is the available power? Is PoE sufficient? Do you need higher voltage (12, 24, 110 V)?
- *Vandal resistance*: How much physical force does the housing have to withstand? Which impact rating (IK class) does that correspond to?
- *Vibrations*: Are there a lot of vibrations in the area, for example, will the camera be placed near rail road tracks or a heavily trafficked freeway? Will it be placed in a bus?

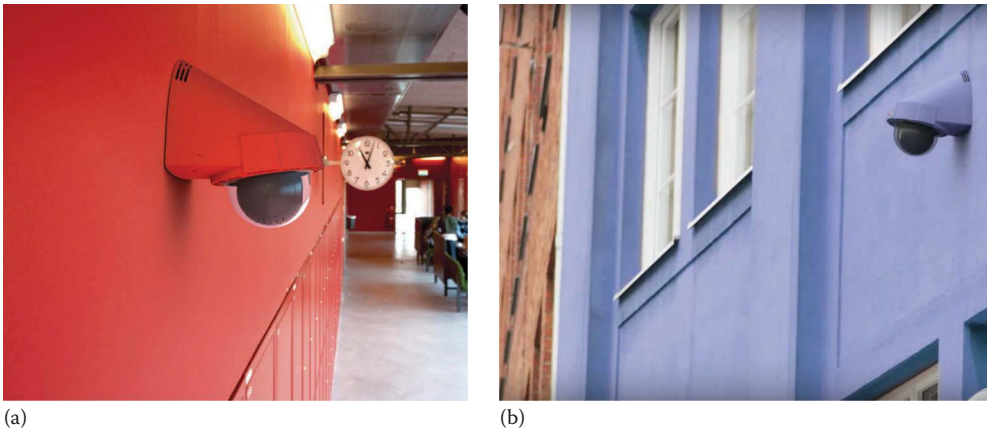


(a)



(b)

**Figure 17.13** An example of housings for fixed cameras (a). Example of a camera mounted on the bottom cover of a housing (b), which is connected to an external power supply that feeds power to the heater and to the camera.



**Figure 17.14** As in these examples, which feature the same PTZ camera model painted in red (a) and cold blue (b), some enclosures can be repainted to blend in with the facade of a building.

And for fixed camera (box camera) housings:

- *Opening*: A side or a slide opening can make a difference when opening the camera for maintenance.

And for dome and PTZ camera housings:

- *Dome*: Do you need a clear, smoked, or mirrored dome?

Some housings also have peripherals such as antennas for amplifying the signal to wireless cameras. An external antenna is only required if the housing is made of metal. Provided that the signal is strong enough, a wireless camera inside a plastic housing usually works without the use of an external antenna. Some vendors offer extra powerful access points that provide the level of connectivity required by wireless network cameras, such as wide coverage and signals that are strong enough to reach the cameras and to ensure the stability of the network.

Sometimes it is preferable that the protective enclosure has the same color as the surroundings, to make the surveillance camera less noticeable or more aesthetically pleasing. Some cameras are designed so that the enclosure can easily be repainted (Figure 17.14).

### 17.3.2 Transparent coverings

To see out from a housing, the camera lens needs a window. Usually, it is made of high-quality glass or durable polycarbonate plastic. Because windows<sup>®</sup> act like optical lenses, they need to be of high quality to minimize their effect on image quality. Imperfections that compromise the clarity of the window material result in lower-quality images.

Windows of housings for PTZ and dome cameras have to meet higher demands. Not only must the windows be specially shaped in the form of a dome, but they must also have high clarity, or imperfections such as dirt particles may be magnified. When the zoom factor is large, even the tiniest particles become problematic. Although smoked domes enable a more discreet installation than clear domes, keep in mind that they also reduce the amount of light available to the camera and therefore have an effect on the camera's light sensitivity. See Figure 17.15 for examples of clear and smoked domes. Some manufacturers also offer mirrored domes. A smoked or mirrored dome acts much like sunglasses do in reducing the amount of light that can pass through the covering. Therefore, the camera might have problems with the depth of field and with adjusting to f-stop changes. To remedy some of those issues, it is best to use auto-iris lenses. With a fixed dome camera, you can use a dome that is partially smoked and partially clear. This makes it more difficult to see where the camera is pointing, but because the lens looks through a clear window, image quality is not compromised.





**Figure 17.15** Examples of clear (a), smoked (b) partially smoked (c) domes for PTZ and fixed dome cameras. Although the smoked dome makes it difficult to see in which direction the camera is pointing, it also reduces the amount of light and therefore the image quality.

### 17.3.2.1 Overcoming the limitations of conventional domes

Camera vendors have long tried to solve the shortcomings of the dome manufacturing process. Until recently, molding technology made it impossible to make a dome in one piece without introducing flaws. To make a dome with the required level of clarity, manufacturers must join two parts: a half-sphere and a cylinder. This joining of parts always results in a transition. The transition is a problem as it affects part of the camera's view, resulting in blurry images. Therefore, the discussion has been limited to where to place the transition and whether it should be smooth and wide (large, slightly blurred area) or sharp and thin (small, very blurred area). When monitoring areas with differences in altitude, such as escalators, hilly roads, or steep arena stands, the transition is particularly troublesome because the camera cannot see clearly above its horizon. The more the camera tilts, the blurrier the image. This phenomenon is called mirroring (see Figure 17.19).

As explained earlier, a window or a dome acts as an extra lens. With conventional domes, the varying distance between the camera and the dome usually causes issues with reflections, distortions, and other optical effects. This is another motivation for trying to solve the manufacturing challenges and spawn a new generation of domes.

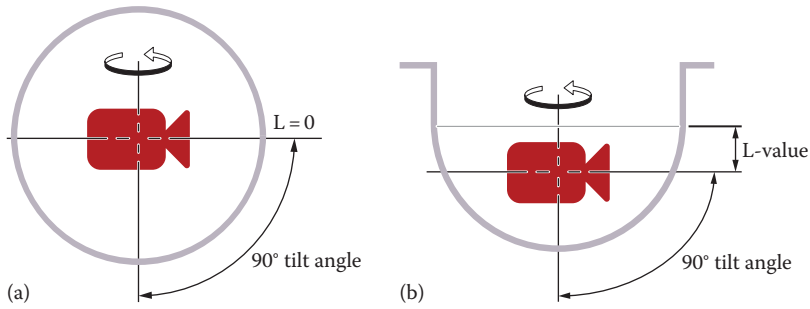
Imagine that a camera's dome is a sphere rather than an elongated half-sphere. To get the best possible image quality, the camera block should be placed at the center point of the sphere, that is, at the zero point on all the axes. However, in most cases the camera is placed lower on the vertical axis to avoid the error of refraction caused by the transition of the dome. The degree of vertical misalignment is called the L-value (see Figure 17.16). This gives the camera a greater tilt range, but also a drop in image quality.

Today, there are PTZ cameras on the market that solve the challenge of the L-value. They consist of two half-spheres that have been joined together and tilted at an angle (see Figures 17.17 and 17.18).

The tilt makes it possible to place the camera block at the center point of the sphere, where the L-value is 0. Because of its position, the camera block remains consistently at the same distance from the dome wall. This means that refractions and other optical effects are kept to a minimum.

Thanks to its elaborate mechanics, this new dome rotates with the camera block. This ability makes it possible to keep an optimal image quality in all pan and tilt positions and to identify with certainty objects as much as 20° above the camera horizon (see Figure 17.19). In other words, cameras with

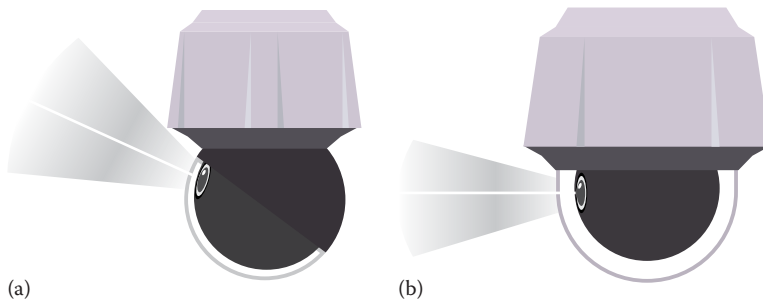




**Figure 17.16** The optimal L-value is zero (a). In most domes, the L-value ends up being well over that (b).



**Figure 17.17** A pan, tilt, and zoom camera with the new type of spherical dome.

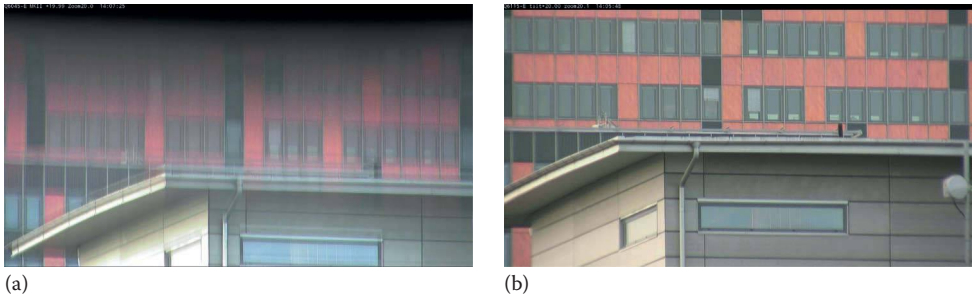


**Figure 17.18** An illustration of the difference between a spherical dome (a) and a conventional dome (b).

this type of dome are much more suited for monitoring uneven terrain than cameras with conventional domes.

What also makes this new dome unique is its ability to shake itself off when it becomes wet, helping the camera to produce sharp images even in rainy weather (Figure 17.20).

When it rains or when someone cleans the dome with a hose, the dome starts rotating at high speed in alternating directions (see Figure 17.21). This breaks the surface tension of the water and the drops fall from the dome.



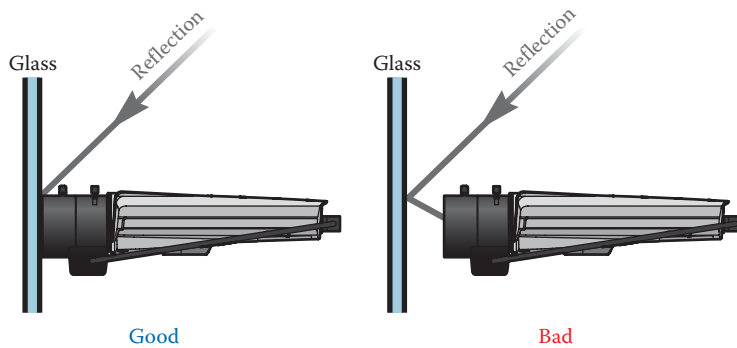
**Figure 17.19** The transition in a conventional dome causes a mirroring effect in the image (a). With the new spherical dome, the image is significantly sharper (b). Both images were taken at a 20° tilt angle and 20 x zoom (a).



**Figure 17.20** Two merged snapshots of the same rainy scene, before shaking off the water (left) and after shaking off the water (right).



**Figure 17.21** An illustration showing how the drops fall off the dome when it shakes.



**Figure 17.22** When installing a camera behind glass, correct positioning of the camera becomes important to avoid reflections.

### 17.3.3 Positioning of fixed cameras

When installing a fixed camera in an enclosure, it is important to position the lens of the camera right up against the window to prevent glare in the image, such as reflections from the camera and the background. To reduce reflection, special coatings can be applied on glass used in front of the lens (see Figure 17.22).

### 17.3.4 Environmental protection

The main environmental threats to a camera—particularly one installed outdoors—are cold, heat, water, and dust.

Housings with built-in heaters and fans (blowers) can handle environments with low and high temperatures. Enclosures that have active cooling with a separate heat exchanger help cameras cope with hot environments.

To withstand water and dust, housings are carefully sealed. In situations where cameras might be exposed to acids, such as in the food industry, housings made of stainless steel are required (see Figure 17.23). Some specialized housings can be pressurized, submersible, bullet proof, or explosion proof. Special enclosures also may be required for aesthetic considerations.



**Figure 17.23** A stainless steel pan, tilt, and zoom camera.

Other environmental elements include wind and traffic. To minimize vibrations, particularly on pole-mounted camera installations, the housing should be small and securely mounted. Electronic image stabilization is an important feature for those installations (see Chapter 3).

The terms *indoor housing* and *outdoor housing* often refer to the level of environmental protection. An indoor housing is used primarily to prevent the entry of dust and does not include a heater or fan. The terms are misleading because the location—whether indoors or outdoors—does not always correspond to the conditions at an installation site. For example, an indoor freezer room, a room with sprinklers, or a parking garage, which is usually cleaned with more aggressive methods such as high-pressure washers, needs a camera that can handle condensing humidity and that has a heater.

### 17.3.5 Vandal and tampering protection

In some surveillance applications, cameras are at risk of violent attacks. Transportation vehicles, schools, prisons, and retail environments are just some examples of areas where vandals or criminals may try to tamper with, redirect, destroy, spray paint, or remove cameras. For examples of cameras for vehicles and prisons, see Figures 17.24 and 17.25.

Although a camera or housing can never guarantee protection from destructive behavior in every situation, there are a number of measures available that can help security managers deal with



**Figure 17.24** An example of a compact, fixed dome camera specially designed for installations in mass-transit vehicles such as buses and trains. It can withstand vibrations, dust, high humidity, and fluctuating temperatures.



**Figure 17.25** A corner-mounted antilitter camera.

camera vandalism. Considerations to keep in mind include camera and housing design, mounting, placement, and intelligent video alarms.

Antiligature cameras are made to resist almost any attempt at violence, whether the intent is self-harm, harming of others, or destruction of the camera or the facilities. Many antiligature cameras fit snugly into corners or have rounded edges so that there are no gripping surfaces (see Figure 17.25). Typically, you find these cameras in correctional facilities and psychiatric wards or in small areas such as interview rooms and elevators.

#### 17.3.5.1 Goals of vandal protection

Many features and best practices can be implemented to increase protection against vandalism. The important goals of vandal protection, regardless of actual technical implementation, include the following:

- *Making it difficult*: Tampering with a video surveillance camera should be difficult. Perhaps even more important is that their design and placement should make them look like they are difficult to tamper with. A vandal should want to think twice before trying to interfere with a camera.
- *Creating uncertainty*: If vandals decide to attack a camera, they should be uncertain as to whether they actually succeeded in destroying the camera or interrupting the recording.
- *Prolonging and delaying*: Even if it is not possible to protect a camera from a determined attack, it is worthwhile to make it very time consuming for a vandal to redirect or destroy the camera. Every second gained increases the chance of discovery or that the vandal gives up.
- *Detecting and sending alarms*: A camera with built-in analytics can detect that someone is tampering with its operation and notify operators. This allows operators to quickly alert field staff to clean, adjust, or replace the camera, to stop the vandal from finishing the attack, or in some other way to fix the problem.

#### 17.3.5.2 Mechanical design

Casings and related components made of metal provide better vandal protection than ones made of plastic. The shape of the housing or camera also matters. A housing or a fixed camera that protrudes from a wall or ceiling is more vulnerable to physical violence than housings or casings for a dome camera, which are often more discreet. A dome's smooth, rounded surface makes blocking the camera's view more difficult. For example, trying to hang a piece of clothing over the camera is nearly impossible. The more a housing or camera blends into an environment, the better its protection against vandalism. For examples of vandal-resistant cameras, see Figures 17.26 through 17.28.

For improved vandal resistance, domes can be made of a durable, transparent material such as polycarbonate plastic, which is the material used to create bullet-proof glass. Increasing the thickness of a dome improves its ability to withstand heavy blows. However, the thicker the dome, the higher the risk of clarity flaws such as small particles embedded in the material. At high zoom levels, these imperfections can be magnified and make the image blurry. Increasing the thickness also can create unwanted reflections and refraction of light, which have a negative impact on image quality.

Special coatings can also be applied to the bodies of cameras and housings to minimize the impact of graffiti. Dirt and drops of water on the dome or housing window cause image distortion. Therefore, some housings have built-in wipers that keep them clean. Some PTZ cameras can vibrate and shake to get rid of water drops from the dome (see Section 17.3.5.1).

#### 17.3.5.3 Mounting

The way cameras and housings are mounted also affects the level of protection. A network camera that is mounted so that all of it is accessible from the exterior is vulnerable to attacks. It is more



**Figure 17.26** Examples of fixed camera housings. The bottom and top left are both classified as vandal-resistant housings.



**Figure 17.27** Examples of fixed dome cameras. All are vandal resistant.

exposed than a camera that uses a recess mount where only the transparent part of the camera or housing is visible (Figure 17.29).

One thing to consider is if it is worth sacrificing the protection that tamper-resistant fasteners offer for the flexibility of standard fasteners. Screws that are not part of standard toolsets can make it more challenging for unauthorized people to dismount cameras and housings from walls and ceilings. The more unusual the screws, the better protection they provide. However, all authorized tasks that involve mounting, dismounting, or moving the cameras become more difficult and expensive because staff need special tools.

When making plans for mounting cameras and protecting the system from vandals, always include the cable runs. Running the cable directly through the wall or ceiling behind the camera provides the best level of protection. A metal conduit is also a good alternative when trying to protect cables from attack.

#### 17.3.5.4 Camera placement

You can also deter vandals by placing cameras in out-of-reach places. A camera that is mounted high up on a wall or in the ceiling is less likely to attract a spur-of-the-moment attack. The downside may be the field of view, which to some extent can be compensated for by selecting a different lens.





**Figure 17.28** A vandal-resistant PTZ camera.



(a)



(b)

**Figure 17.29** An example of a PTZ camera with a recessed mount. When the camera is mounted, you can see the dome and the trim ring (a) but not the actual mounting bracket (b).

#### 17.3.5.5 Intelligent video protecting cameras

Analytics in network cameras and VMSs can help protect cameras against vandalism. Intelligent algorithms can detect if a camera has been redirected, obscured, or tampered with in other ways and can send alarms to operators in central control rooms or to staff in the field. These types of algorithms include detecting if the view changes (tampering), if there are abnormal sounds (audio), and if the camera is subjected to violence (shock).

Without this type of intelligence, keeping track of the proper functioning of hundreds of cameras in demanding environments is too difficult. In systems where no one is actively watching live video, analytics simplifies automatic surveillance by notifying staff when someone interferes with a camera's operation. For more information about analytics and their applications, see Chapters 15 and 16.

### 17.3.6 Mounting types

Because the need for surveillance is not limited to a specific type of space, the variety in mounting options must be vast. To minimize vibrations, always make sure that the camera mount is stable. Because PTZ cameras move around, the action can cause image interference if the camera mount is not properly secured. In outdoor situations, sturdy mounting equipment is necessary to avoid vibrations caused by strong winds. If the mount is not strong or stable enough, the worst-case scenario is that the camera falls and damages people or property (Figure 17.30).

#### 17.3.6.1 Ceiling mounts

Ceiling mounts are used primarily in indoor installations. The enclosure itself can be the following:

- *Surface mount*: Mounted directly at the surface of a ceiling or wall and therefore completely visible. This mount is also known as a hard-ceiling mount.
- *Flush mount*: Mounted inside the ceiling with only parts of a camera and housing (usually the dome) visible. This mount is also known as a recessed mount or drop-ceiling mount.
- *Pendant mount*: Hung from a ceiling like a pendant.
- *Covert mount*: Typically used to mount tiny modular cameras in spaces where discretion is key. These mounts can be completely covert, barely visible (pinhole mount), partly visible (flush mount), or fully visible (surface mount).

Figure 17.31 provides examples of each mounting type.

#### 17.3.6.2 Wall mounts

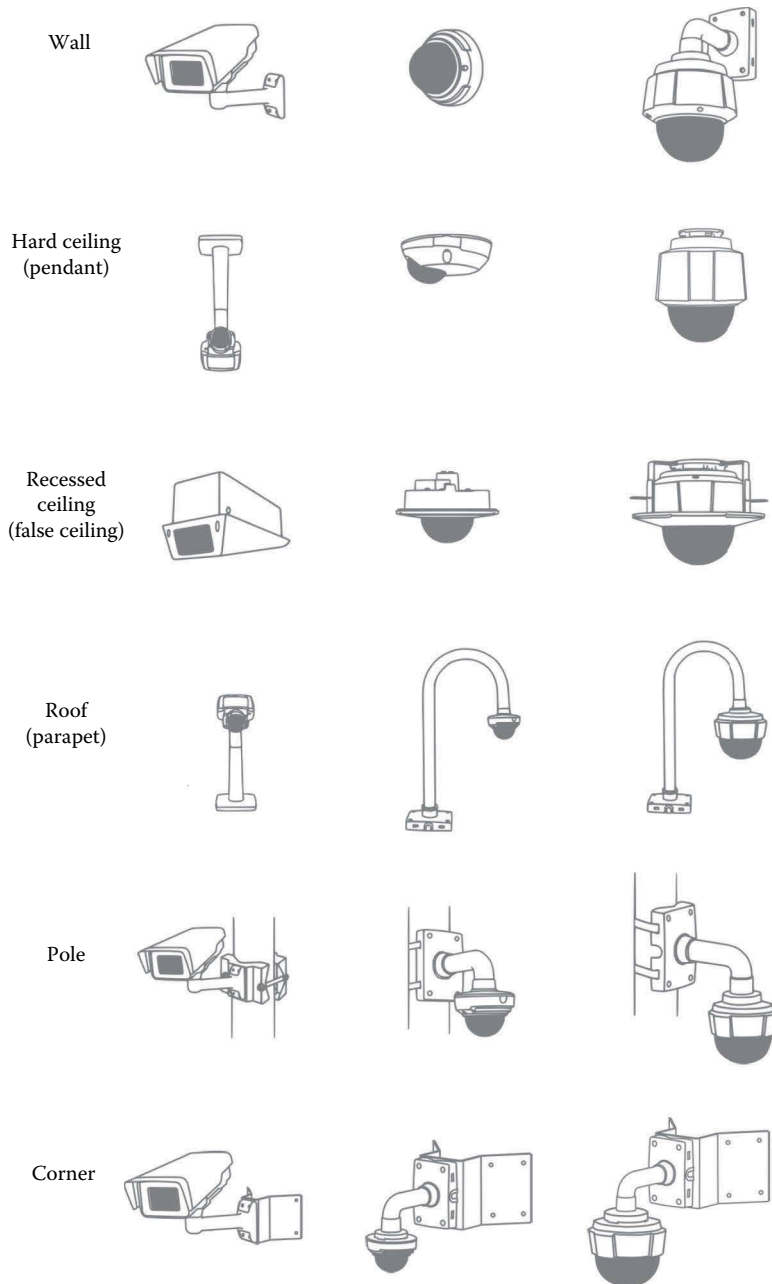
Wall mounts are often used to mount cameras inside or outside a building. The cable can often be routed through the arm, and many mounts have an internal cable gland or gasket to protect the cable (Figure 17.32).

#### 17.3.6.3 Pole mounts

Pole mounts (Figure 17.33) often hold PTZ cameras in large outdoor areas such as parking lots, roads, and city squares. This type of mount is usually designed to minimize the effects of wind and ground vibrations. The pole and the mount must be able to absorb these vibrations and limit their impact on the camera. When calculating the sway, consider the height and diameter of the pole, as well as the material. Concrete poles sway less than metal and wooden poles. Factor in the weight and dimensions of the equipment that the pole needs to carry. This is especially important for PTZ cameras and cameras with high optical zoom. If the pole is underdimensioned, you risk an unreasonable amount of motion blur in the images. More advanced PTZ cameras have built-in electronic image stabilization to limit the effect of wind and vibrations. However, heavy cameras can cause serious injuries if they fall down. As with wall mounts, the cable can usually run inside the pole, and cable exits and outlets must be sealed properly.

#### 17.3.6.4 Parapet mounts

Parapet mounts (Figure 17.34) are used to mount cameras on rooftops or to raise the camera for a better angle of view. A benefit of parapet mounts is that the camera is cheaper and easier to service than if it is hung from a wall mount. Because the arm can swing inward, maintenance staff can access the camera from the rooftop rather than having to use a lift or another type of aerial work platform.

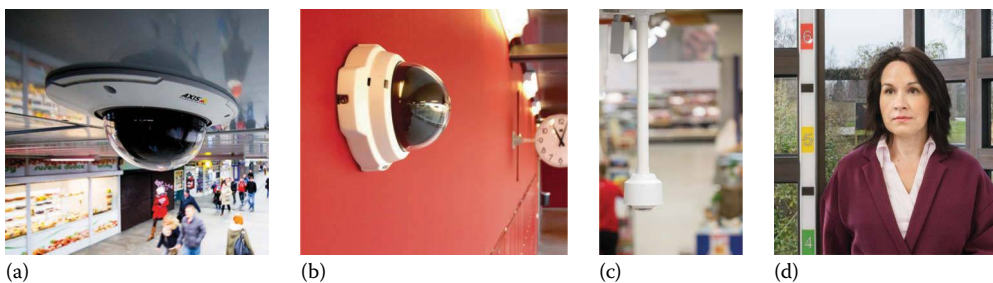


**Figure 17.30** Examples of common mounting types.

#### 17.3.6.5 Special mounts

Some spaces require special mounting solutions. When a camera needs to hang from the corner of a building, a corner adapter on a standard wall mount provides the solution (Figure 17.35).

A telescopic pendant mount allows a camera to hang low from a tall ceiling. If it has a ball joint, the pendant mount can hang from a sloped ceiling regardless of the angle or swing if hit by forklifts or tall slow-moving vehicles (Figure 17.36). Other special mounts include pipe adapters for mounting cameras on standard threaded pipes and recess mounts for mounting cameras in soffits. Some mounts work with both PTZ cameras and fixed dome cameras, although they sometimes require adapter kits.



**Figure 17.31** An example of a flush mount (a), a surface mount (b), a pendant mount (c), and a covert mount (the camera is hidden in the height strip) (d).



**Figure 17.32** Examples of a corner wall mount (a) and a regular wall mount (b).



**Figure 17.33** An example of a pole-mounted PTZ camera.



**Figure 17.34** An example of parapet-mounted cameras on a rooftop.



**Figure 17.35** A corner adapter with a pendant mount kit for a fixed dome camera. See also Figure 17.32 of a corner wall-mounted installation.

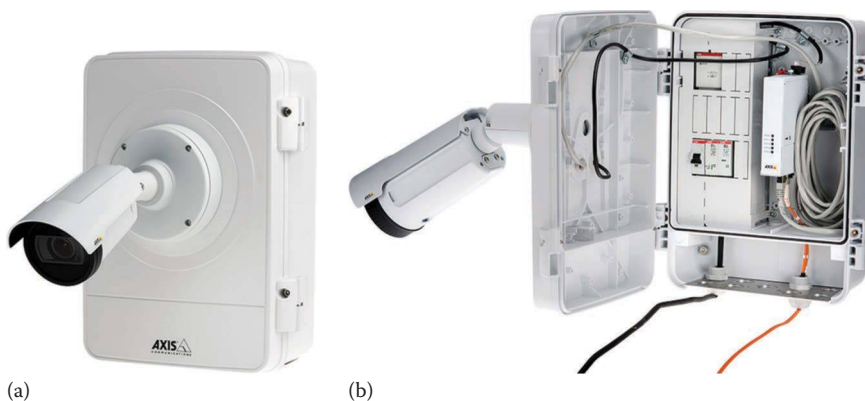
Many outdoor camera installations require a cabinet or connection box for peripherals such as power supplies and media converters (for connection with fiber-optic cables). Rather than a standard connection box, which tends to taint the visual appearance, you can use a cabinet with a built-in camera mount, which is more aesthetically pleasing and easier to install (see Figure 17.37). Such cabinets can cater to special requirements on data communication systems, such as cell modems, wireless routers, and switches, including modular attachment options such as DIN rails and clips.

Sometimes, the connections need to be contained within a junction box. Junction boxes make the connections neater and practical because they are accessible behind the front panel, whereas their concealment protects them from tampering. More importantly, they cage sparks and heat from loose connections and short circuits. Some camera vendors can supply a complete solution, including a wide range of modular mounting systems, cabinets, and conduit adapters, as well as back boxes and junction-box plates that fit both the gangs of standard junction boxes and the hole patterns of the camera mounts. For examples, see Figure 17.38.





**Figure 17.36** A telescopic pendant mount that swings if pushed.



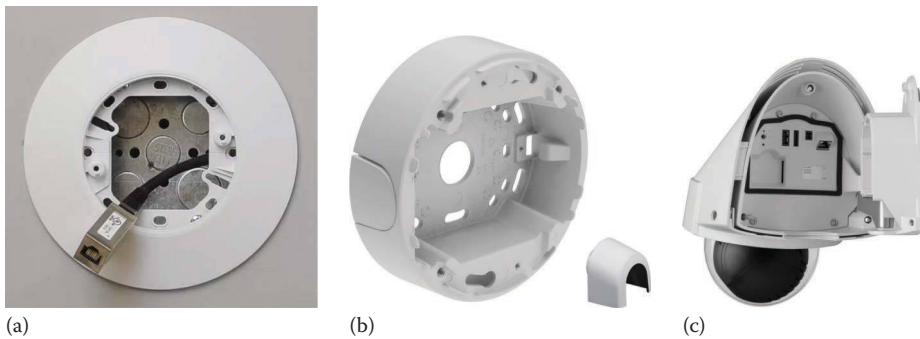
**Figure 17.37** A cabinet with a bullet-style camera mounted on it (a). Inside, there is room for media converters, surge protectors, power supplies, switches, and other electrical peripherals (b).

Pan-tilt motors bring remote pan- and tilt functionality to fixed cameras. The motor itself is usually mounted on a pole or a wall. Sometimes the pan-tilt motor is the base for a dual camera (a thermal and a conventional camera in one housing) and IR illuminators; see Figure 17.39.

### 17.3.7 EMC

All electric and electronic devices release electromagnetic energy, also known as radio-frequency (RF) emission. Emission is a by-product of electrical or magnetic activity. Unfortunately, the emissions from one device can interfere with other devices, which can lead to data loss and picture quality degradation on monitors and cameras and cause other equipment to malfunction. When they say that a device has electromagnetic interference or electromagnetic compatibility (EMC) problems, they mean that the device cannot withstand emissions from other devices or that its own emissions are so severe that the device can interfere with other devices.





**Figure 17.38** Examples of junction boxes and accessories: a junction-box plate that fits a single gang box, double gang box, 4-in. square junction box, or 4-in. octagon junction box. This plate is used to mount a small fixed dome camera (a); a conduit back box that can be mounted on a hard, flat surface or a junction box. This back box is used to mount a fixed dome camera and fits a  $\frac{3}{4}$ -in. conduit or a conduit adapter such as the one shown beside the back box (b); a PTZ camera with a built-in wall mount and junction box (c).



**Figure 17.39** An example of a pan-tilt motor with a dual camera and IR illuminators.

The other main concept of EMC is immunity. Immunity is a measure of how good a system is at rejecting interference from other devices. The opposite of immunity is susceptibility. It refers to the tendency of a product to malfunction or break down when exposed to emissions.

Before determining how a product shall be installed in an existing environment, always take both emission and immunity into consideration.

#### 17.3.7.1 EMC standards

All network video manufacturers must declare the EMC of their network video products. In most of the world, this includes emission as well as immunity.

In Europe, EMC is included in the CE mark (see Figure 17.41), which in turn is included in the EU's harmonization legislation. Conformance with the EMC directive is mandatory for network

cameras, but manufacturers can self-certify their products. Usually, they show compliance through the following standards:

- *Emission*: EN 55022, and sometimes EN 61000-6-3 or EN 61000-6-4
- *Immunity*: EN 55024, EN 61000-6-1, and EN 61000-6-2

The standards that begin with EN 55022 and EN 55024 are product standards that are valid specifically for information technology equipment (ITE). They are harmonized with international standards (CISPR 22) and, for example, the corresponding standards in Canada (ICES-003) as well as Australia and New Zealand (AS/NZS CISPR 22). The EN 61000-6 standards are generic standards.

In the United States, the Federal Communications Commission (FCC) stipulates the rules and regulations for telecommunication devices. Network cameras are included in the set of rules called CFR 47, Title 47, Part 15 Radio-Frequency Devices, Subpart B—Unintentional Radiators. However, the FCC rules only refer to emission and not immunity.

Harmonic current emissions and flicker are covered by separate EU standards (EN 61000-3-2 and EN 61000-3-3). Moreover, there are additional standards for specific types of digital products or applications. For example, railway applications should follow EN 50121-4 and IEC 62236-4, audio and audiovisual equipment EN 55130-4, and alarm systems EN 50130-4. There are also standards for wireless equipment.

Because IT, audio, and visual equipment is getting harder to distinguish (you can use your TV to access the internet and your computer to watch TV), the different standards are about to merge. From October 2013, manufacturers can use the new standard EN 55032 (CISPR 32) to show EMC compliance. The scope of this standard covers IT, audio, and video equipment, which are grouped into the term *multimedia equipment* (MME). The standard committees are also working on an immunity standard for MME, EN 55035 (CISPR 35).

### 17.3.7.2 Emission

Emission refers to the ability of equipment to function satisfactorily without emitting too much electromagnetic energy that can disturb other equipment in that environment.

When manufacturers declare a product's compliance with emission requirements, they state which environmental category (Class A or Class B) the product can be used in. Although the limits are not identical in all regions, the relevant standards use the same categories:

- *Class A digital devices*: Intended for commercial, industrial, or business environments. Due to the higher level of RF interference already present in commercial markets, the emissions requirements are less stringent than Class B devices.
- *Class B digital devices*: Intended for residential areas. Because these products are used in home environments, they have to comply with stricter emission requirements than Class A devices.

### 17.3.7.3 Immunity

Immunity is a measure of the ability of electronic products to tolerate the influence of electromagnetic phenomena and electrical energy (radiated or conducted) from other electronic products. For example, any system connected to the alternating current mains power line must be immune to transient surges. The required immunity level depends on the type of appliance and which environment it is intended to operate in.

The requirements for a product to withstand electrical disturbances or interference are the opposite of emission requirements. That is, the immunity requirements are higher in an outdoor, industrial, or similar environment than in a residential environment.

#### 17.3.7.4 Choosing between shielded and unshielded network cables

In the United States, Class A devices can be connected with unshielded twisted-pair (UTP) network cables. This is because the FCC rules disregard immunity and, as mentioned earlier, the emission limits are less stringent for Class A devices than for Class B devices.

When products have been tested against both the FCC Class A and the FCC Class B requirements, you can choose between the following:

- Use unshielded twisted-pair (UTP) network cables and fulfill the FCC Class A limits for emission.
- Use shielded twisted-pair (STP) network cables and fulfill the FCC Class B limits for emission.

In demanding electrical environments, it is best to use a STP network cable. Outdoor environments always fall into that category, so always use an STP cable if the camera is used outdoors. One might think that the term *demanding electrical environments* only applies to outdoor environments. This is not true. Even indoor environments can be electrically demanding. Examples of such environments are installations where the network cable runs parallel with electrical mains supply cables or where large inductive loads such as motors or contactors are close to the camera or its cable. It is also mandatory to use an STP cable with an indoor camera or encoder if the cable is partially or fully routed outdoors (see Section 17.3.8).

In summary, in the United States most indoor surveillance deployments can use UTP cables and still fulfill the regulatory requirements. For the customer, UTP network cables are attractive because they are cheaper, are less bulky, and allow simpler termination of the connector.

In Europe and most other countries, the limits for immunity require that both Class A and Class B devices be connected with STP cables.

For more information about unshielded and shielded cables, see Chapter 9.

### 17.3.8 Safety of electrical equipment

The low-voltage directive (LVD) provides broad objectives for the safety of electrical equipment. Its purpose is to align the certification of electrical products so that they are safe to use in all EU member states. Just like the EMC directive, LVD is included in the CE mark, and manufacturers can self-certify their products.

For ITE, compliance with LVD is usually shown through the EN 60950 standard and its international equivalent, IEC 60950. IEC stands for the International Electrotechnical Commission and is an organization that publishes standards for all electrical, electronic, and related technologies. In the United States, Occupational Safety and Health Administration (OSHA) has given Underwriters Laboratories (UL) the right to perform certification of ITE. This standard is known as UL 60950.

The standards' identical numbers indicate that they are harmonized, which means that in essential areas they are equal. However, there are regional differences. Therefore, products that are marketed and sold on multiple markets need to show all the required certifications. In the United States, UL is recognized by OSHA, within its Nationally Recognized Testing Laboratory program, to perform certification of ITE. The standard is known as UL 60950. The purpose is to ensure that the products meet the requirements of both the construction and general industry OSHA electrical standards. In Canada, the equivalent certification body is the Canadian Standards Association, but usually the European or American certifications are sufficient for this market.

The parts that apply to network cameras, encoders, and their power supplies are IEC/EN/UL 60950-1 and for outdoor products also IEC/EN/UL 60950-22. Products with built-in LEDs also need to comply with EN 62471, which among other things includes exposure limits to prevent hazard to eyes and skin.

In 2014, a new hazard-based standard was published. EN 62368-1 replaces EN 60950, but until the latter is withdrawn in 2019 the two standards will coexist. In parallel, IEC and UL have been developing sister standards with the same number.

As mentioned earlier, the purpose of the safety standards is to ensure that products are safe to use without risk of personal injury or property damage caused by hazards such as electric shock, fire, dangerous temperatures, and mechanical instability.

In outdoor or demanding electrical environments or even simply if the cables are routed outdoors, devices must be protected from power surges. There are several ways to ensure that a power surge has a path to reach ground. One way is to use power sourcing equipment (whether it is a midspan, an endspan, a network switch, a power supply, or any other end device) that is properly grounded and to use STP network cables to connect all devices. If the camera has a grounding screw, both ends of the grounding wire must be in contact with their respective grounding surfaces. For power supplies, the safety standards include a specific set of conditions, such as limited power source and safety extralow voltage (SELV). Simply put, that a power supply fulfills the requirements for SELV means that it is built in such a way that its voltage stays within safe values.

An installation should follow the safety standard and the national electrical code both to protect the investment (limiting the risk of damage through power surges) and to make the electrician's job easier, faster, and cheaper.

### 17.3.9 Environmental ratings

Various standards groups have defined different classes of protection, so there are a number of different environmental classifications that manufacturers can label their products with. The most common ratings are IP, NEMA, and IK, which are explained in the following. A brief description of the European ATEX certification also is provided.

#### 17.3.9.1 IP ratings

International protection (more commonly known as IP) ratings define the level of protection that electrical appliances provide against the intrusion of solid objects or dust, accidental contact, and water. So it is not surprising that IP is often assumed to stand for ingress protection. The ratings are based on the harmonized international and European standard IEC/EN 60529. Just like the safety standards mentioned in the section about safety of electrical equipment, the IP standard is connected to the LVD, and manufacturers can self-certify their products. Although companies have to make sure that they fulfill the basic requirements of the standard, some subject their products to tougher tests than others. For example, an IP test is more difficult to pass if the product first had to undergo an impact test.

Figure 17.40 shows an example of an IP66-rated enclosure for video encoders. It allows a non-IP-rated video encoder to be installed outdoors and still be protected against the elements. Such



**Figure 17.40** An example of an IP66-rated protective enclosure for video encoders.

enclosures are available for cameras too (see Section 17.3), although in new installations it is usually more cost efficient to use outdoor-ready cameras.

An IP rating consists of the letters IP followed by two digits and sometimes two more letters. The first digit indicates the level of protection that an enclosure provides against access to hazardous parts and ingress of solid foreign objects. Examples of hazardous parts are electrical conductors or moving parts. Examples of solid objects are fingers, tools, or dust. The higher the number, the better the protection. The second digit indicates the level of protection against intrusion by liquids. Again, the higher the number, the better the protection. For example, an IP66 rating means it is dust tight and protects against ingress from powerful water jets. Products intended for outdoor use should have an IP rating of at least IP44, although most outdoor installations require IP66. When there is no protection rating given with regard to either solid objects or liquids, the letter X is used (e.g., *IP2X*). See Tables 17.3 and 17.4 for explanations of the first two digits. Because the final two letters are optional and rarely used, they are not included in the tables.

If the second digit is 6 or lower, this implies compliance also with the requirements for all levels below it. So an IP65 enclosure is automatically approved for environments that demand IP55 or IP64 levels of protection. However, equipment that is designated with a second digit of 7 or 8 should

**Table 17.3** IP ratings, first digit: Foreign solid objects

Level	Protected against	Effective against
0	Not protected	No protection.
1	Objects larger than 50 mm	A large surface of the body such as back of the hand, but no protection against deliberate contact with a body part.
2	Objects larger than 12.5 mm	Fingers or other objects can penetrate as far as 80 mm as long as it is safe from hazardous parts. Objects with a diameter of 12.5 mm cannot penetrate fully.
3	Objects larger than 2.5 mm	Objects, such as tools and thick wires, cannot penetrate at all.
4	Objects larger than 1 mm	Objects, such as wires and screws, cannot penetrate at all.
5	Dust protected	Ingress of dust is not completely prevented, but dust does not enter in sufficient quantity to interfere with satisfactory operation of the equipment.
6	Dust tight	No ingress of dust.

**Table 17.4** IP ratings, second digit: Liquids

Level	Protected against	Effective against
0	Not protected	No special protection.
1	Dripping water	Dripping water (vertically falling drops) has no harmful effect.
2	Dripping water when tilted up to 15°	Vertically dripping water has no harmful effect when the enclosure is tilted at any angle up to 15° from its normal position.
3	Spraying water	Water falling as spray at an angle up to 60° from the vertical has no harmful effect.
4	Splashing water	Water splashed against the enclosure from any direction has no harmful effect.
5	Water jets	Water projected from a nozzle against the enclosure from any direction has no harmful effect.
6	Powerful water jets	Water from heavy seas or water projected in powerful jets cannot enter the enclosure in harmful quantities.
7	Brief immersion in water	Ingress of water in a harmful quantity cannot be possible when the enclosure is immersed in water under defined conditions of pressure and time.
8	Continuous submersion in water	The equipment is suitable for continuous submersion in water under conditions that shall be specified by the manufacturer. The conditions must be harsher than for IPX7 (see previous).
9	Water from high pressure and steam jet cleaning	Water directed at the housing from any angle under very high pressure has no harmful effect.

not be used where it might be exposed to water jets. Therefore, never assume that an IP67-rated product can also withstand IP66 environments unless it is dual coded, that is, IP66/IP67.

### 17.3.9.2 NEMA ratings

The National Electrical Manufacturers Association (NEMA) is a U.S.-based association that provides standards for electrical equipment enclosures. NEMA has adopted and published a harmonizing IP standard, ANSI/IEC 60529 through the American National Standards Institute (ANSI). However, they also have their own standard, NEMA 250, which they have launched successfully on the global market.

Like the IP standard, NEMA 250 addresses ingress protection, but it also considers other items such as corrosion resistance, performance, and construction details (see Tables 17.5 and 17.6). Therefore, it is safe to say that a NEMA type is comparable to an IP rating, but it would be wrong to state the opposite. These table cannot be used to convert IP ratings to NEMA types, and the assumption of equivalence should always be verified through testing.

NEMA 250 and the National Electrical Code, also known as NFPA 70, define hazardous (or classified) locations as locations that may contain high enough quantities of hazardous materials, such as gases, vapors, combustible dusts, fibers, or flyings, to create an explosion. See also Section 17.3.13.7. This means that cameras or housings that are used in these environments must be inherently safe, that is, they cannot be the cause of an explosion. In North America, the most common category is Class 1 Division 1 as this is the highest level of protection, and products in this category can be used in just about any explosive environment.

The UL standards for enclosures, UL 50 and UL 50E, are based on the NEMA 250 standards. The major difference between them is that while NEMA allows self-certification, UL enforces compliance by demanding that products pass third-party testing and inspection.

### 17.3.9.3 IK ratings

Many security cameras are placed in environments where they are subjected to various kinds of impacts. The most obvious causes of impact are vandalism and other physical attacks, but falling branches, debris caught in the wind, climbing animals, and resting birds can also cause impact damages. Also, even an experienced installer can drop a camera.

For enclosures, there is a standard that specifies degrees of protection against external mechanical impact. It was originally approved in 1994 by the European Committee for Electrotechnical Standardization as the European standard EN 50102. When it was adopted as an international standard in 2002, it changed numbers to IEC/EN 62262. The standards are identical and both are still valid.

Similar to the IP ratings, the degrees of protection against impact are indicated by a code. It consists of the letters IK followed by two digits (see Table 17.7).

The tests are used to demonstrate an acceptable level of robustness when assessing the safety of a product. The main concern is to make sure that the inner equipment of a product is properly protected by the enclosure. Although the product inside the enclosure needs to be safe from accidental or intentional probing after impact, it does not necessarily have to be operational. During the test, each exposed surface is hit five times and evenly distributed over the surface. The same point, or area around it, cannot be hit more than three times. The points to which impact should be applied is specified in the relevant product standard. The product standard may also include exceptions to the rule of maximum five hits. After the test, the product needs to be evaluated: Are the damages admissible? Is the product still safe and reliable? So really, an IK10 rating does not mean that the product is resistant to impact but is instead more a measure of robustness.

Manufacturers who are serious about providing high-quality products go further than the requirements of the standard. To ensure that the product keeps its level of robustness throughout its lifespan, they may test the weakest part of the camera rather than the strongest. As mentioned in Section 17.3.9.1, they may extend their efforts by doing IP tests after they have performed IK tests. Others may check



**Table 17.5** NEMA ratings for enclosures in nonhazardous locations

NEMA	IP	Indoor	Outdoor	Protected against
Type 1	IP10	•		Access to hazardous parts and ingress of solid foreign objects (falling dirt). No protection against liquids.
Type 2	IP11	•		Access to hazardous parts and ingress of solid foreign objects (falling dirt). Ingress of water (dripping and light splashing).
Type 3	IP54	•	•	Access to hazardous parts and ingress of solid foreign objects (falling dirt and windblown dust). Ingress of water (rain, sleet, snow). Will be undamaged by the external formation of ice on the enclosure.
Type 3R	IP14	•	•	Access to hazardous parts and ingress of solid foreign objects (falling dirt). Ingress of water (rain, sleet, snow). Will be undamaged by the external formation of ice on the enclosure.
Type 3S	IP54	•	•	Access to hazardous parts and ingress of solid foreign objects (falling dirt and windblown dust). Ingress of water (rain, sleet, snow). The external mechanisms remain operable when ice laden.
Type 3X		•	•	Access to hazardous parts and ingress of solid foreign objects (falling dirt and windblown dust). Ingress of water (rain, sleet, snow). Provides an additional level of protection against corrosion. Will be undamaged by the external formation of ice on the enclosure.
Type 3RX		•	•	Fulfills the requirements of NEMA 3R and NEMA 3X.
Type 3SX		•	•	Fulfills the requirements of NEMA 3S and NEMA 3X.
Type 4	IP56	•	•	Access to hazardous parts and ingress of solid foreign objects (falling dirt and windblown dust). Ingress of water (rain, sleet, snow, splashing water, and hose-directed water). Will be undamaged by the external formation of ice on the enclosure.
NEMA 4X	IP56	•	•	Access to hazardous parts and ingress of solid foreign objects (windblown dust). Ingress of water (rain, sleet, snow, splashing water, and hose-directed water). Provides an additional level of protection against corrosion. Will be undamaged by the external formation of ice on the enclosure.
Type 5	IP52	•		Access to hazardous parts and ingress of solid foreign objects (falling dirt and settling airborne dust, lint, fibers, and flyings). Ingress of water (dripping and light splashing).
Type 6	IP67	•	•	Access to hazardous parts and ingress of solid foreign objects (falling dirt). Ingress of water (hose-directed water and the entry of water during occasional temporary submersion at a limited depth). Will be undamaged by the external formation of ice on the enclosure.
Type 6P	IP67	•	•	Access to hazardous parts and against ingress of solid foreign objects (falling dirt). Ingress of water (hose-directed water and the entry of water during prolonged submersion at a limited depth). Provides an additional level of protection against corrosion. Will be undamaged by the external formation of ice on the enclosure.
Type 12	IP52	•		Without knockouts. Access to hazardous parts and ingress of solid foreign objects (falling dirt and circulating dust, lint, fibers, and flyings). Ingress of water (dripping and light splashing).
Type 12K	IP52	•		With knockouts. Access to hazardous parts and ingress of solid foreign objects (falling dirt and circulating dust, lint, fibers, and flyings). Ingress of water (dripping and light splashing).
Type 13	IP54	•		Access to hazardous parts and ingress of solid foreign objects (falling dirt and circulating dust, lint, fibers, and flyings). Ingress of water (dripping and light splashing). Spraying, splashing, and seepage of oil and noncorrosive coolants.

**Table 17.6** NEMA ratings for enclosures in hazardous locations

NEMA	IP	Indoor	Outdoor	Constructed for use
Type 7		•		In hazardous (classified) locations classified as Class I, Division 1, Groups A, B, C, or D as defined in NFPA 70
Type 8		•	•	In hazardous (classified) locations classified as Class I, Division 1, Groups A, B, C, and D, as defined in NFPA 70
Type 9		•		In hazardous (classified) locations classified as Class II, Division 1, Groups E, F, or G, as defined in NFPA 70
Type 10				Meets the requirements of the Mine Safety and Health Administration, 30 CFR, Part 18

**Table 17.7** IK ratings

Level	IK01	IK02	IK03	IK04	IK05	IK06	IK07	IK08	IK09	IK10	IK10+
Impact energy (Joule)	0.14	0.2	0.35	0.5	0.7	1	2	5	10	20	50+ <sup>a</sup>
Mass (kg)			0.2			0.5		1.7	5		
Drop height (mm)	56	80	140	200	280	400	400	300	200	400	

<sup>a</sup> IEC/EN 62262 provides for a maximum resistance of IK10 at 20 J with the possibility of extending the impact energy up to 50 J. Some types of equipment need more protection. Therefore, the market has extended the test beyond what the standard provides. These IK ratings are known as IK10+. When using such a rating, the manufacturer should indicate the impact energy, mass, and drop height of the striking element.

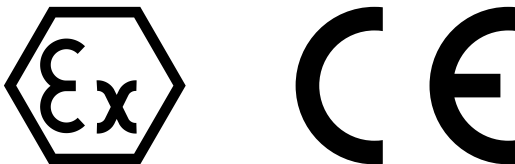
that the product is still operational after the IK test. This does not imply that it retains its IP rating, although a well-engineered product could be assumed to tolerate such an environment.

#### 17.3.9.4 IECEx and ATEX certifications

When a camera is installed in a potentially explosive environment, its housing must meet very specific safety standards. The international standard is known as IECEx or *International Electrotechnical Commission System for Certification to Standards relating to Equipment for use in Explosive Atmospheres*. In Europe, products intended for these environments must comply with the ATEX directive, which stands for *appareils destinés à être utilisés en atmosphères explosibles* (or, in English, *equipment for potentially explosive atmospheres*). ATEX-certified products (see Figure 17.42) bear the hexagon-shaped Ex mark and CE mark (see Figure 17.41). In North America, the NEMA Class 1 Division 1 rating is generally preferred over the ATEX and IECEx certifications.

An explosive atmosphere is defined as an area where flammable substances, such as liquids, vapors, gases, or combustible dusts, are likely to occur and mix with air in such proportions that excessive heat or sparks might make them explode. Examples of such areas include gas stations, oil platforms and refineries, chemical processing plants, printing industries, gas pipelines and distribution centers, grain handling and storage, aircraft refueling and hangars, and hospital operating theaters.

It is the explosive environment that must be protected from potential igniters from the camera and other equipment. In other words, the camera must be explosion protected, explosion proof, or flame proof. These terms are used synonymously and mean that the camera will contain any explosion originating within its housing and will not generate conditions that could ignite vapors, gases, dust, or fibers in the surrounding air. It does not mean that the camera itself will withstand an exterior explosion (Figure 17.42).



**Figure 17.41** ATEX and CE marks.



**Figure 17.42** An example of a PTZ camera designed and certified for use in potentially explosive atmospheres and harsh environmental conditions. Because the housing is made of stainless steel, it is suitable for offshore, onshore, marine, and heavy industrial environments.

**Table 17.8** Zone divisions of explosive atmospheres

Zone (IECEX and ATEX)	Atmosphere	Definition	Class and division (NFPA 70)
Zone 0	Gas	An area in which an explosive mixture is continuously present or present for long periods	Class I Division 1
Zone 1	Gas	An area in which an explosive mixture is likely to occur in normal operation	Class I Division 1
Zone 2	Gas	An area in which an explosive mixture is not likely to occur in normal operation and if it occurs, it will exist only for a short time	Class I Division 2
Zone 20	Dust	An area in which an explosive mixture is continuously present or present for long periods	Class II Division 1
Zone 21	Dust	An area in which an explosive mixture is likely to occur in normal operation	Class II Division 1
Zone 22	Dust	An area in which an explosive mixture is not likely to occur in normal operation and if it occurs, it will exist only for a short time	Class II Division 2

Becoming ATEX- or IECEx-certified is a complicated process. Unlike most of the other European standards and directives mentioned in this chapter, manufacturers cannot usually self-certify their explosion-protected products. They must have them tested and certified by a notified body (NB), which is appointed by the member state, or an Ex Certification Body (ExCB), which is appointed by IECEx. ATEX has an NB exception for gas and dust atmospheres that are classified as low-risk atmospheres; that is, areas in which an explosive mixture is not likely to occur in normal operation, and, if it does occur, will exist only for a short time.

Explosive atmospheres are divided into categories, gas and dust, where each category is divided into zones (see Table 17.8). There are also two categories of equipment, one for mining (I) and one for surface industries (II).

### 17.4 STORAGE AND SERVER CONSIDERATIONS

Depending on the system size and requirements, setting up the video management part of a network video system can be anything from a trivial 5-minute task to a very complex and time-consuming activity. Designing a server and storage system begins with a few basic decisions: Whether to use a central or distributed architecture, what level of performance the recording servers must have,

and how much storage is needed? The required performance of the servers, possible system architecture, and suitable storage setups also depend on which video management software you use. For more information on storage and servers, see Chapter 12.

### 17.4.1 Small system: From 1 to 10 cameras

A small system usually consists of a few cameras using edge-based recording to SD cards or a NAS drive. Some installations use network video recorders (NVRs) with or without built-in PoE ports. A laptop, a smartphone, or tablet can be used to monitor the video (see Figure 17.43).

### 17.4.2 Midsize system: From 10 to 100 cameras

A typical midsize installation has a server that runs VMS software. Additional storage is attached to it (Figure 17.44), but a smaller system may use an NVR instead. To increase performance and reliability, the storage is usually configured with RAID. The video is normally viewed and managed from a client rather than from the recording server.

### 17.4.3 Large system: From 100 to 1000+ cameras

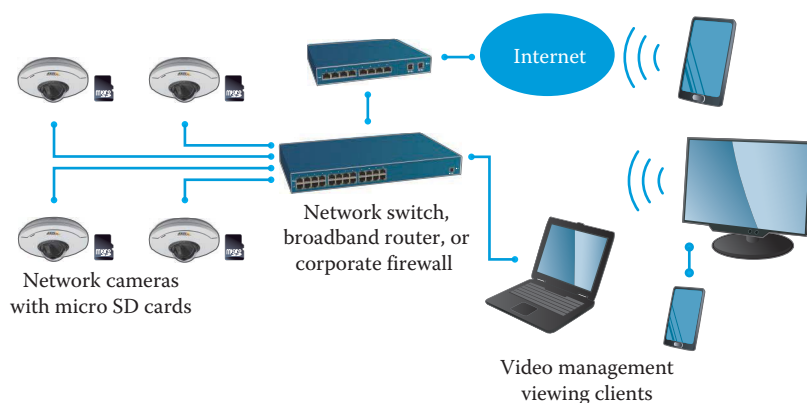
To manage large amounts of data and bandwidth, a large-sized installation requires multiple high-performing and reliable servers dedicated to video management tasks (Figure 17.45). A setup with dedicated storage servers allows you to balance the load and, when needed, scale up the system by adding more storage servers or perform maintenance without disrupting the whole system.

### 17.4.4 Federated systems

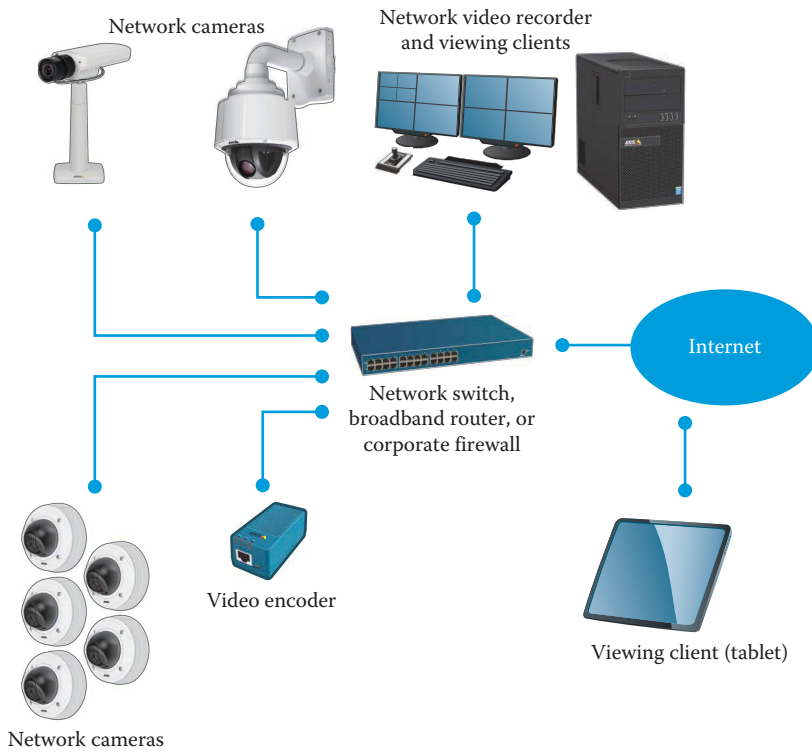
In enterprise installations, such as a global company, a large hospital, or a school district with many buildings, there are several video surveillance systems located at different facilities. Some of those systems might be small, some mid-sized, and some large. In a true federated enterprise system, the subsystems are all tied together into one system. This is known as system federation. This means that all cameras and systems can be managed, monitored, and maintained from one location (Figure 17.46). Each site records and stores the video from local cameras.

### 17.4.5 Provisioning the server

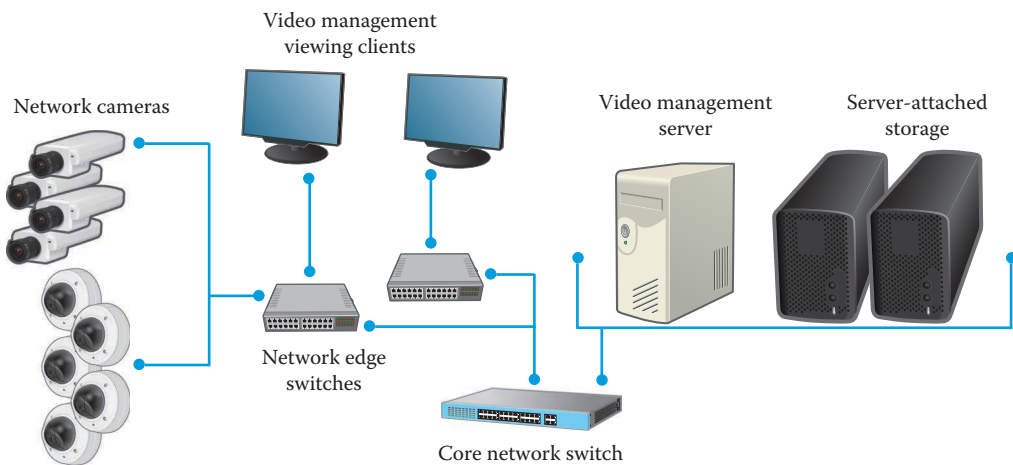
A PC server used for video management should be properly provisioned to handle the current camera system and have the ability to scale up if needed in the future. Each server has a certain baseline of how many cameras it can handle based on its central processing unit, network card, internal random access memory, and hard drives. You also have to factor in the total number of cameras, their resolution and frame rate, and the retention goal of the system.



**Figure 17.43** A small system with network cameras, SD cards, and multiple viewing devices.

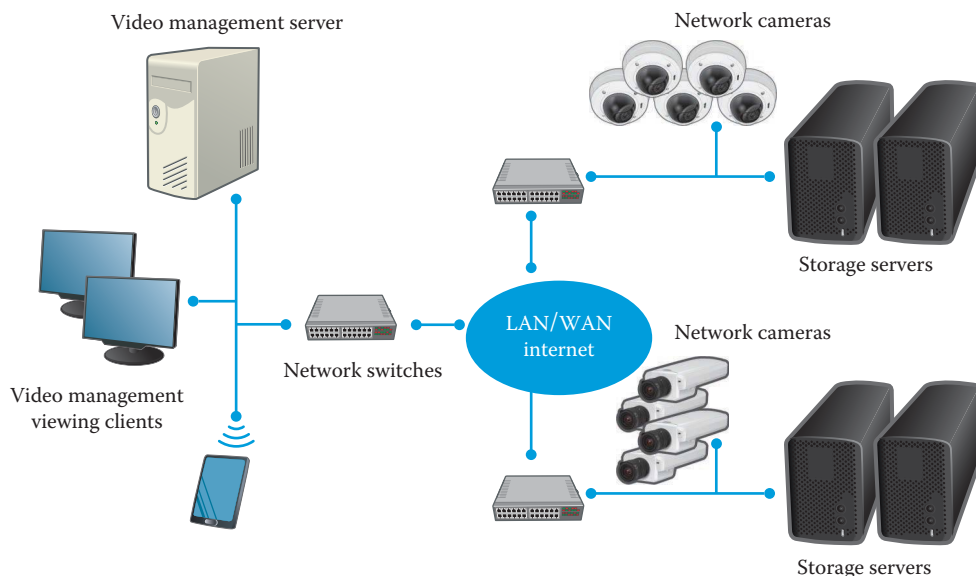


**Figure 17.44** A midsize system.

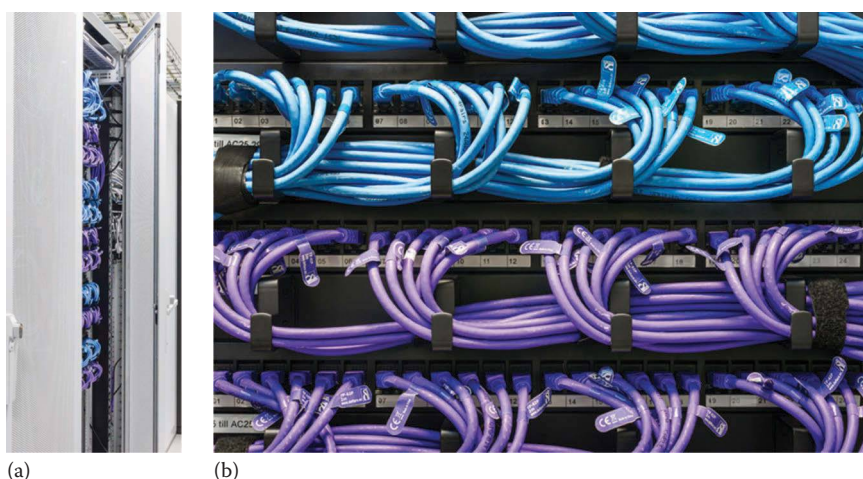


**Figure 17.45** A large system.

Depending on the task and how much load it puts on the server, some physical servers only use a fraction of the processing power and working memory. To save on hardware, reduce rackspace, and limit power consumption, many set up virtual servers. This means that one server can support several systems and applications. But it can also mean that several physical servers are merged into one more powerful machine to handle an increased load. A dedicated server, which runs one application, is always faster than a server running multiple applications.



**Figure 17.46** A distributed system that is also federated into one system.



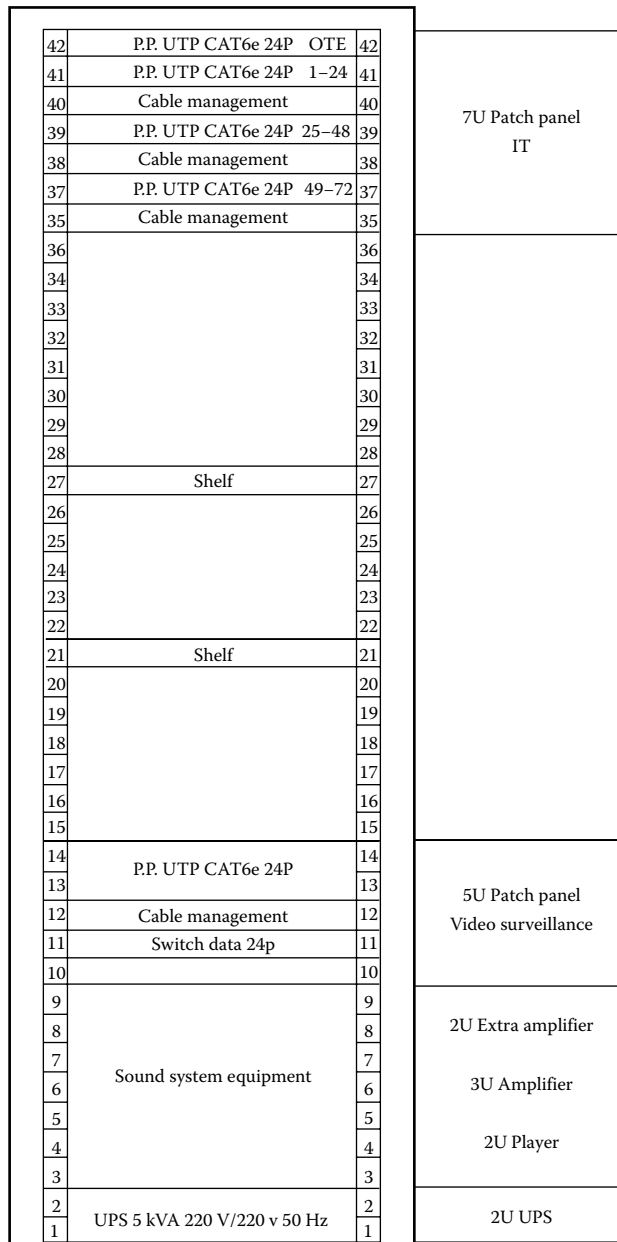
**Figure 17.47** An example of a server rack (a) and a close-up of a patch panel with network cables (b).

The server racks should be well organized (see Figure 17.47). Separate the surveillance system from other systems such as POS systems. Do not share cables between the surveillance system, the POS system, or other data equipment. Figure 17.48 shows an example of a rack layout used in a retail installation.

### 17.4.6 Calculating storage

Calculating the appropriate amount of storage is a very important task when designing a video surveillance system. However, it is not an exact science because the size of video files depends on the complexity and amount of motion in a scene. Some guidelines, along with an example of how the amount of storage can be limited by changing some parameters in the system, are provided in the following.





**Figure 17.48** An example of a rack layout.

#### 17.4.6.1 Calculating storage needs

Proper calculation of the storage required in a network video system is essential for the success of the network video project. The factors to consider when calculating storage needs include

- The number of cameras
- The number of hours per day the camera will be recording
- How long the data must be stored, also known as retention time
- Event-triggered recordings only or continuous recording
- Other parameters such as frame rate, compression, image quality, and complexity

**Table 17.9** Example of storage requirements calculation for H.264

Resolution	Frames per second	Bitrate (Mbit/s)	GB/hour	Hours of operation/day	GB/day
4CIF	5	0.569	0.26	8	2.1
	12	1.07	0.48	8	3.8
	24	1.65	0.74	8	5.9
	30	1.88	0.85	8	6.8
HDTV 720p	5	1.70	0.77	8	6.2
	12	3.23	1.45	8	11.6
	24	4.93	2.22	8	17.8
	30	5.61	2.52	8	20.2
HDTV 1080p	5	3.82	1.72	8	13.8
	12	7.28	3.28	8	26.2
	24	11.1	5.00	8	40.0
	30	12.6	5.67	8	45.4

**Table 17.10** Example of storage requirements calculation for motion JPEG

Resolution	Image size (kB)	Frames per second	Bitrate (Mbit/s)	GB/hour	Hours of operation/day	GB/day
4CIF	46	5	1.84	0.83	8	6.63
	46	12	4.42	1.99	8	15.92
	46	24	8.83	3.97	8	31.76
	46	30	11.04	4.97	8	39.76
HDTV 720p	132	5	5.28	2.38	8	19.04
	132	12	12.67	5.70	8	45.60
	132	24	25.34	11.40	8	91.20
	132	30	31.68	14.26	8	114.08
HDTV 1080p	298	5	11.92	5.36	8	42.88
	298	12	28.61	12.87	8	102.96
	298	24	57.22	25.75	8	206.00
	298	30	71.52	32.18	8	257.44

One of the factors affecting storage requirements is video compression. Sample storage calculations for two compression formats, H.264 and Motion JPEG, are given in Tables 17.9 and 17.10. Because of the number of variables that affect average bitrate levels, calculations are not so clear-cut for H.264. With Motion JPEG, there is a clear formula because the video file consists of several individual image files. Storage requirements for Motion JPEG recordings vary depending on the frame rate, resolution, and level of compression.

The amount of motion in a scene can have a big impact on the amount of storage required. The following numbers are based on continuous recording with lots of motion in a scene, such as a train station. With fewer changes in a scene, the figures can be significantly lower.

H.264 calculations:

$$\frac{\text{Approximate bitrate}}{8} \times 3600\text{s} = \frac{\text{MB per hour}}{1000} = \text{GB per hour}$$

$$\text{GB per hour} \times \text{Hours of operation per day} = \text{GB per day}$$

$$\text{GB per day} \times \text{Retention time in days} = \text{Storage need}$$

Motion JPEG calculations:

$$\text{Image size} \times \text{Frames per second} \times 3,600 \text{ s} = \frac{\text{KB per hour}}{1,000,000} = \text{GB per hour}$$

$$\text{GB per hour} \times \text{Hours of operation per day} = \text{GB per day}$$

$$\text{GB per day} \times \text{Retention time in days} = \text{Storage need}$$

Note that these calculations are examples only and do not take into consideration any overhead or other technical issues that could result in greater file sizes. Nor do they take into consideration storage space for the operating system or video management software. Also consider that formatting the drive reduces the amount of available storage. For example, a 2 TB drive might yield 1.82 TB of available memory after formatting.

## 17.5 PROVISIONING NETWORK BANDWIDTH

When designing a network video system, it is important to properly design the network and associated bandwidth. As bandwidth was scarce, network design was a big challenge in the early days of network video. Today's networks, where we count in the tens or hundreds of gigabits, make it less challenging because, if designed correctly, they can easily cope with today's large amounts of network video. Network video products use network bandwidth based on their configuration. Bandwidth usage, as with storage, depends on image resolution, compression, and frame rate, as well as the complexity of the scene.

### 17.5.1 Limiting the bandwidth

Today's gigabit networks can easily cope with the demands of network video systems. There are also several technologies available that enable the management of bandwidth consumption:

- *Switched networks and virtual local network (VLANs)*: Using VLANs on a switched network—a common networking technique today—the same physical computer and video surveillance network can be separated into two autonomous networks. Although these networks remain physically connected, the network switch logically divides them into two virtual and independent networks. Fiber interconnections between switches are ideal for systems with multiple sites where intermediate distribution frames (IDFs) and main distribution frames (MDFs) can interconnect through a fiber backbone.
- *Quality of service (QoS)*: Using QoS will guarantee a certain bandwidth to a specific application or to a certain camera in a video surveillance system.
- *Multicast*: With multicast you can reduce bandwidth on networks where multiple clients request a video stream. Multicast requires that all switches, routers, and other networking equipment support multicast as well.
- *Event-driven frame rate*: A rate of up to 25 or 30 fps on all cameras at all times is above the level required for many applications. With the configuration capabilities and built-in analytics of network cameras and video encoders, frame rates under normal conditions can be set lower—for example, 5 fps—which will dramatically decrease bandwidth consumption. In the event of an alarm (e.g., if motion detection is triggered), the recording frame rate can be increased automatically. In many cases, it is sufficient to have the camera send video over a network only if the video is worth recording; the rest of the time nothing needs to be sent.

### 17.5.2 Network and system latency

In a video surveillance system, it may be important to consider latency. This is especially true when the video is being monitored live and when a PTZ camera is in operation. A well-designed network should have very little latency, typically in the hundred-millisecond range. In a network with many hops, latency can be upward of one second or more, which can present problems.

More important in a surveillance system is to measure the total end-to-end system latency, from live image to viewing on the monitor. Factors that may affect latency include shutter speed, encoding, cable infrastructure, network, video management software, server hardware, client hardware, and decoding.

### 17.5.3 Network cabling

Never underestimate the importance of the cables in a wired network. Poorly or incorrectly installed network cables can cause numerous problems. Even the smallest cabling issue can have serious effects on the operation of the network. A kink in a cable can cause a camera to respond intermittently, and a poorly crimped connector may prevent PoE from functioning properly. Consider the following when installing cables:

- Use the correct wiring standards.
- Do not combine the wiring standards T568a and T568b on the same cable.
- Use high-quality CAT6 cables.

Cables are categorized according to the data rates that they can transmit effectively. The specifications also describe the material, the connectors, and the number of times each pair is twisted per meter. Nowadays, the most used cable type is CAT5e or better. Make sure that the cabling you use match the requirements of the installation.

- CAT3 with 16 MHz bandwidth (no longer used)
- CAT5e with 100 MHz bandwidth
- CAT6 up to 250 MHz
- CAT6a up to 500 MHz
- CAT7 up to 600 MHz
- CAT 7a with a frequency range through 1000 MHz

For gigabit connectivity and a future-proof installation, even if your existing network switches and routers only support 100 Mbps, use CAT6 cabling or beyond. This way, you avoid having to purchase and reinstall the cabling infrastructure when the bandwidth requirements increase.

Cabling for the video surveillance system is usually done by construction department contractors. This cabling work should include:

- Pulling the cables in the ceiling
- Installing racks and patch panels
- Terminating (patching the cables) and labeling them
- Making sure that at least 30 cm (12 in.) of cable hangs in the ceiling from each point in the drawing where cameras shall be installed
- Certifying of the cables to warrant their work

#### 17.5.3.1 Tips for better network cabling

- The maximum network cable run between devices is 100 m (325 ft).
- If using sockets, take the distance between the socket and the computer into account. A good rule of thumb is 90 m (300 ft) for horizontal runs and 10 m for the patch cabling.
- Use the same connector and cable type, such as STP, for the whole length of cable.
- To avoid interference, keep the network cable runs separate from the electrical mains cabling.
- To avoid fire hazards and violations of building codes, do not suspend network cabling from ceiling tiles.
- Use shielded network cables outdoors to protect products from surges and interference or if required by local regulations (see Sections 7.3.7 and 7.3.8).

- Because network cabling typically uses solid wires, cabling should not be twisted or bent into a tighter radius than four times the diameter of the cable.
- When fastening cable runs, do not use metal staples or adjust cable ties too tightly.
- Avoid using a daisy chain network topology.
- When preparing the cables in the ceiling, pull at least 30 cm (1 ft) extra cable to use as a service loop in case of termination failures that require retermination.

### 17.5.3.2 Preparing the network cable

A network cable consists of four pairs of twisted wires. The wires are color coded (orange, green, blue, and brown). The cable specification has been designed for high-speed data transfer and very little cross talk.

- Do not untwist more than 6 mm ( $\frac{1}{4}$  in.) of the cable at either end. If you do, it could cause problems such as near-end cross talk, which will have a detrimental impact on your network.
- Keep the pairs together.
- Wire the plug correctly, making sure that each wire is properly connected to its respective pin at both ends.
- Use the correct connectors. Network connections usually use RJ45 connectors specifically designed for either stranded or solid cable. Match the connector with the cable type, such as STP or UTP.
- Use the correct crimping tool for the specific type of connector.

### 17.5.3.3 Certifying the cable installation

Installers who need to prove to the network owner that the installation has been done correctly and meets Telecommunications Industry Association (TIA) or International Organization for Standardization (ISO) standards need to certify their work. Network owners who want to guarantee that the infrastructure is capable of handling the video bandwidth can use a tester to certify the network infrastructure. In some cases, testers are employed to pinpoint specific problems. Certification tests are vital if there is any discrepancy between the installer and network owner after an installation has been performed.

In twisted-pair copper wire networks, copper cable certification is achieved through a thorough series of tests in accordance with standards set by the TIA or the ISO. These tests are done using a certification testing tool, which gives a pass or fail indication.

A cable tester is used to verify that all of the intended connections exist and that there are no unintended connections in the cable being tested. When an intended connection is missing, it is “open.” When an unintended connection exists, it is a “short” (a short circuit). If a connection leads to the wrong place, it is “miswired,” meaning that it has two faults: it is open to the correct contact and shorted to an incorrect contact.

Generally, the testing is done in two phases. The first phase, the “opens test,” makes sure each of the intended connections is good. The second phase, the “shorts test,” makes sure there are no unintended connections.

## 17.6 TOOLS FOR DESIGNING SYSTEMS

Designing a network video system requires making a lot of choices and fine-tuning. A modern video surveillance system is complex, includes many components, and is based on best-of-breed technology. Therefore, it is crucial to have tools that make choosing the right components easier. System designers must ask themselves: Which are the right:

- Camera types?
- Enclosures and environmental protection?

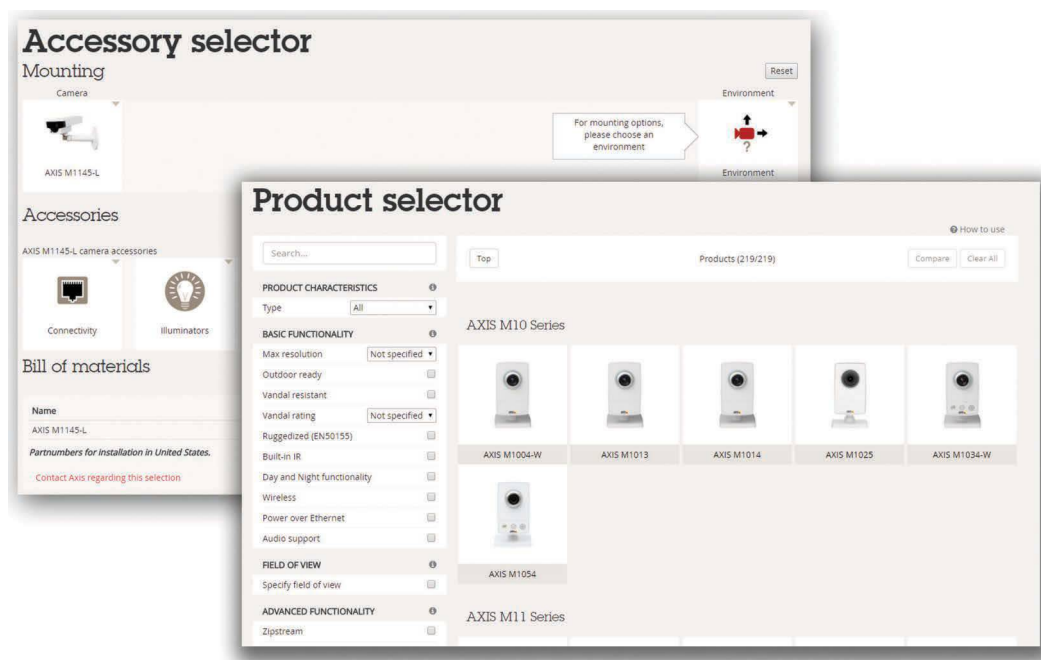
- Lenses to provide the desired field of view?
- Camera locations and coverages?
- Camera mounts and accessories?
- Cabling infrastructures?
- Network infrastructure, bandwidth, and security?
- Servers for running the VMS and connecting the storage?
- Sizes and performances of the storage units?
- Rack designs for the server room?
- Video management software solutions?

There are many different tools available for different stages of the design. Some tools are basic and help you choose a few of the components, while others are comprehensive, advanced tools for total system designs. The following sections cover component selection tools as well as system design tools. Some of the advanced system design tools can use extensions for CAD software, which is covered at the end.

### 17.6.1 Calculators and component selection tools

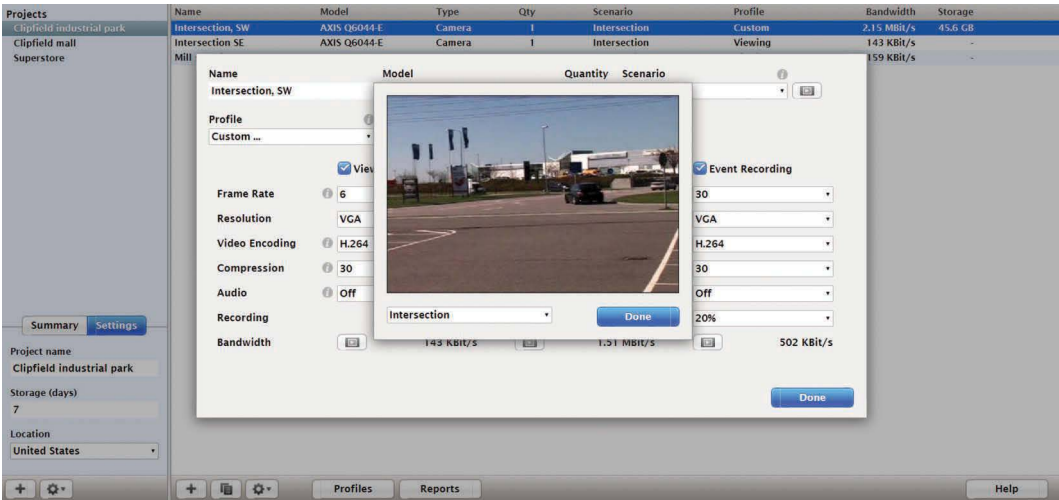
To validate the field of view, there are lens calculators (see Chapter 4). To assist in choosing the correct product for the installation, there are product selection guides available, as well as guides for selecting the right accessory for the camera. Some are available both online (see Figure 17.49) and as smartphone apps.

Other popular tools are those that allow integrators and system designers to estimate the storage and network bandwidth of their customers' systems. These tools reduce the risk of over- or under-specifying the network and storage needs. Based on the performance of each camera and sample videos for a few typical scenarios, the tool delivers a customized bill of materials (BOM) for each project. The sample videos also help users understand the effect of typical image scenarios and what a certain resolution and frame rate mean (see Figure 17.50).



**Figure 17.49** Examples of online product and accessory selectors.



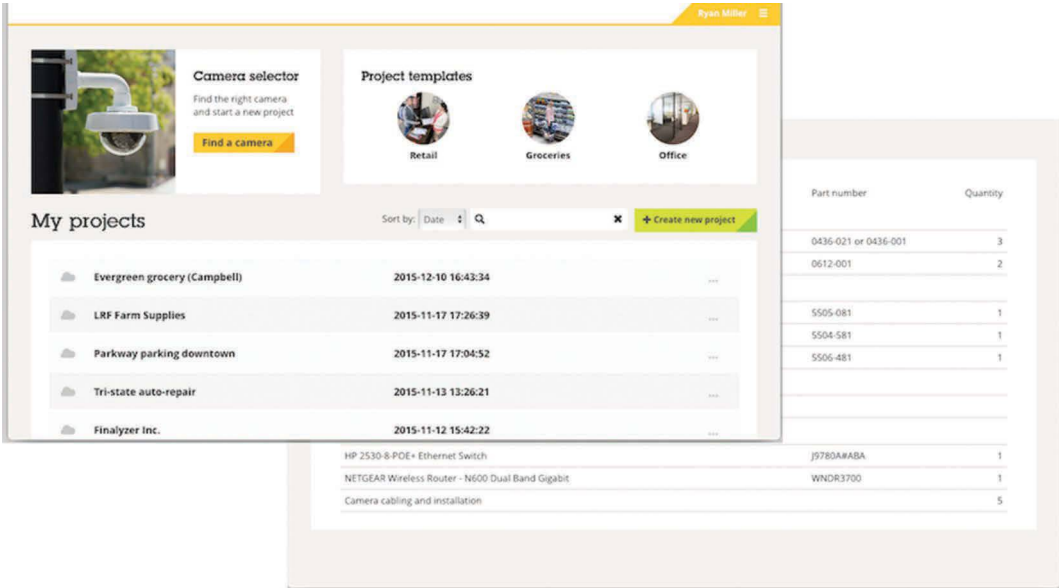


**Figure 17.50** An example of a system design tool for calculating bandwidth and storage requirements and understanding what effect frame rate, resolution, and compression have on the quality of the video.

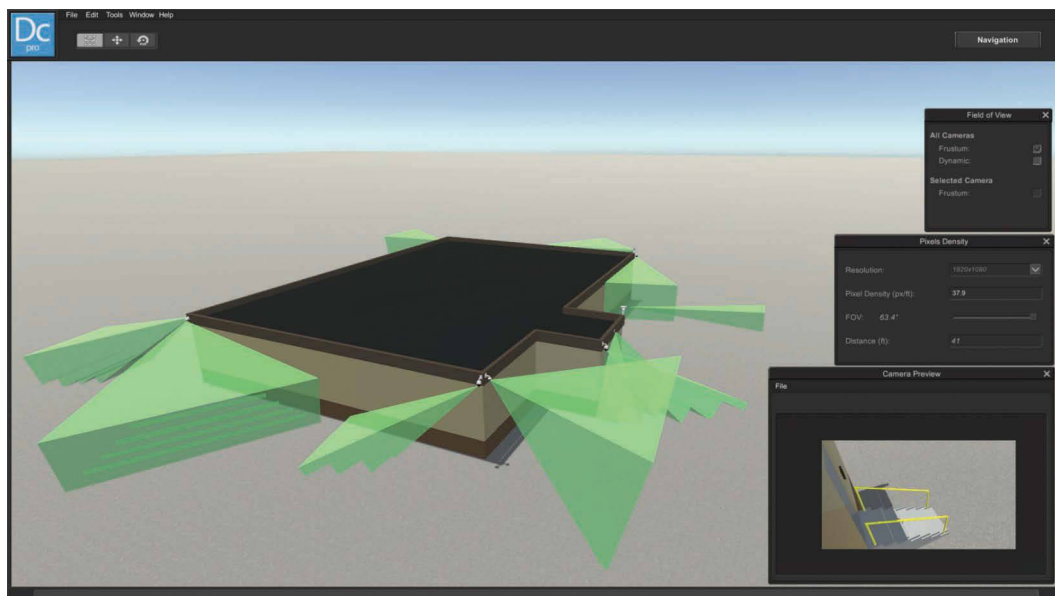
17.6.2 Comprehensive system design tools

Demands for accessible and visually appealing yet inclusive design tools are growing like never before, especially now that we have become accustomed to a world with smart devices and easy-to-use apps. System design tools make picking the right camera and recording solution for any surveillance scenario quick and easy. When the project design is finished, you get a BOM that includes everything you need for the installation (Figure 17.51).

You can keep track of multiple projects, and templates provide a quick way to get new projects started. Set up your scenario and get a shortlist of cameras, mounts, accessories, and recording solutions that meet your needs.



**Figure 17.51** System design tools collect several product selection tools and system calculators into a single intuitive and user-friendly tool.



**Figure 17.52** An advanced system design tool offers a wide range of options and great flexibility. (Image courtesy of Iomnis Surveillance Solutions, Houston TX.)

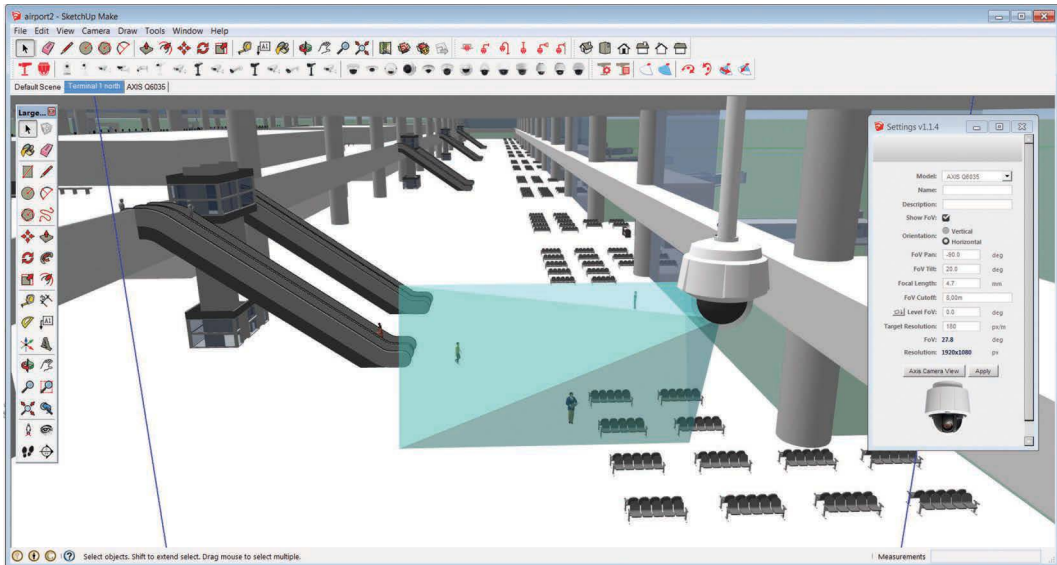
Many types of system architecture benefit from a more comprehensive toolset than what was previously available. Some vendors are now starting to fill this gap by providing very advanced system design packages. These include all of the features in the tools described in this chapter, and they connect the data from the design process. In this way, human error can be reduced and efficiency greatly increased.

Beyond recommending cameras based on the user-defined data, an advanced design tool allows the user to define requirements such as indoor, outdoor, WDR technology, and form factor (see Figure 17.52). It provides 3D virtualization of the cameras and their intended environment, as well as field-of-view representation. Once the camera selections are complete, the program calculates and recommends servers, storage solutions, and network switches. It also provides a BOM, compiled associated data sheets, meantime between failure (MTBF) documents, and a combined CSI-formatted specification based on the final camera list. The platform builds commonly used outputs for requests for proposal (RFP) including floor plans with field-of-view representations and camera legends.

A true comprehensive design tool even includes tools for designing the server rooms (whether MDF or IDF), allows for drag-and-drop placement of appliances in vertical racks, and generates a 3D or 2D design complete with calculations for power.

### 17.6.3 Extensions for CAD software

For building models, there are more advanced tools that plug into 2D or 3D CAD software. Each design company has their own preference in software based on price and required experience. Therefore, some manufacturers supply extensions for a range of different CAD platforms, such as SketchUp®, Autodesk® Revit, and Microsoft® Visio®. Three-dimensional modeling is particularly effective when trying to engage the customer and align their expectations with reality.



**Figure 17.53** An extension tool for SketchUp® makes it easy to visualize the camera's coverage.

### 17.6.3.1 SketchUp®

Extensions for SketchUp saves time and gives system designers reassurance that they are specifying the right type of camera and mounting accessories for the surveillance system. Interactive 3D CAD camera models let you view the camera's coverage in any SketchUp model. Through its integration with Google Maps™, SketchUp can also give a good overview of any outdoor surveillance project.

Place the cameras in the security system designs. Review how the camera fits into the building layout and easily spot if objects such as columns or walls obstruct the camera's view (see Figure 17.53). To determine the optimal setup and the best video surveillance coverage, adjust the field of view through PTZ. If the tool has a view function, you can view the SketchUp model as the camera sees it.

To complete the system design, the extension provides a detailed camera listing and a BOM that include all the products included in the design.

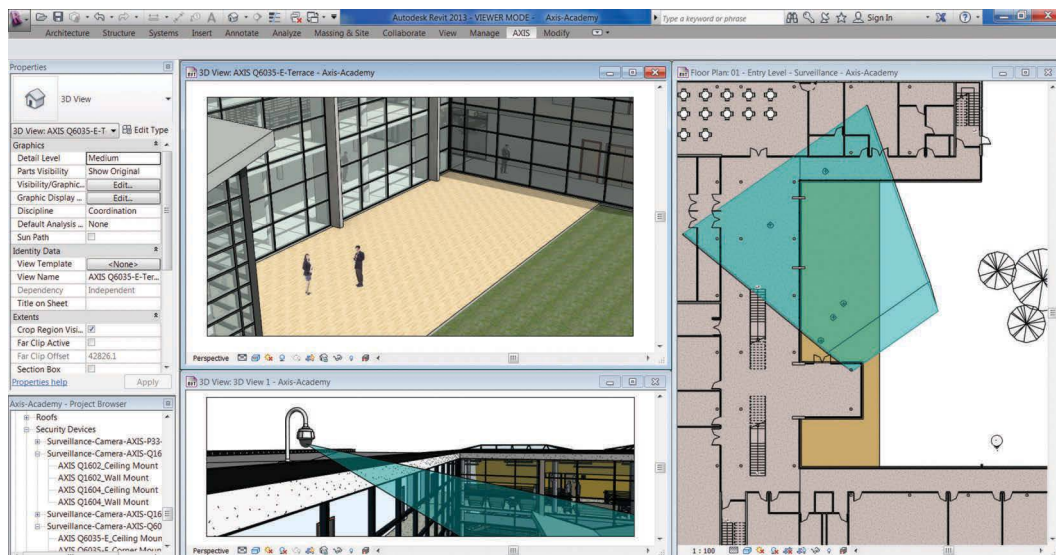
### 17.6.3.2 Revit®

System designers using Autodesk Revit can also install sets of interactive 3D CAD camera models. Like the extension for SketchUp, you can place security cameras directly in the building layout and visualize each camera's coverage, making sure that they cover the critical areas and avoid blind spots (see Figure 17.54). The extension provides the metadata needed for building information modeling, which makes it easy to integrate surveillance system planning into Revit projects.

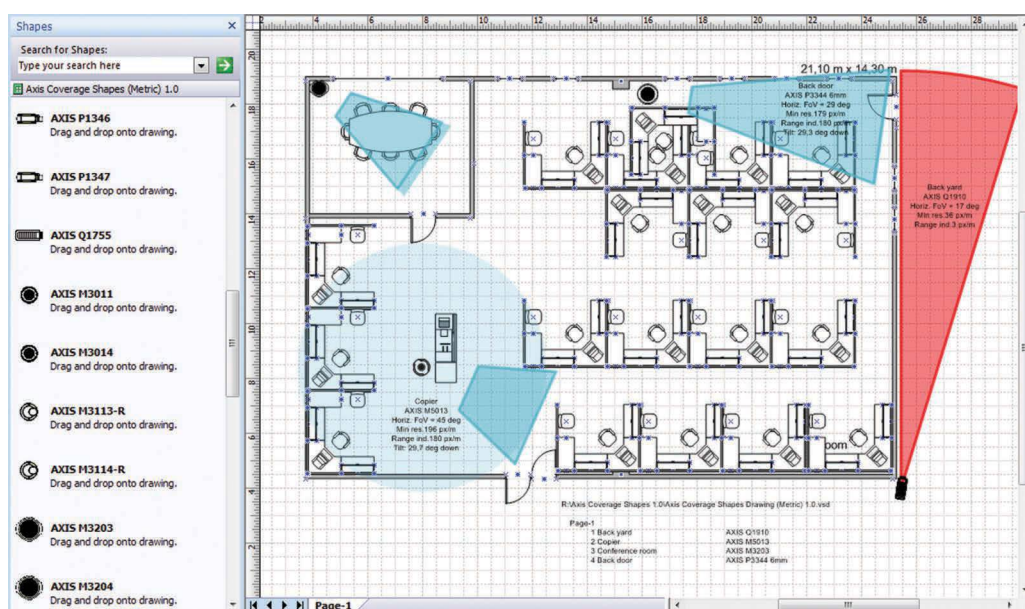
### 17.6.3.3 Visio®

Microsoft Visio libraries of surveillance cameras give system designers the confidence to know that the right number and type of cameras are specified for each surveillance system. Being able to give customers a better understanding of what the final project will look like, even if only in 2D, can also save approval time.

Import an existing drawing into Visio or use the built-in drawing tools to create your floor plan. Then drag cameras from the stencil into the drawing to see the cameras' coverage based on the field of view and resolution (see Figure 17.55).



**Figure 17.54** Installing 3D CAD models of cameras makes it easy to integrate video surveillance into a building project.



**Figure 17.55** Use Visio® stencils to get visual representation of the cameras' coverage.

## 17.7 LEGAL ASPECTS

Video and audio surveillance can be restricted or prohibited by laws. Each region or country has its own sets of rules that regulate video surveillance. For example, notifying the public of the existence of video surveillance is usually mandatory. These notifications do not have to be intrusive, but they should nevertheless be visible (see Figure 17.56). Always check the local laws before installing a video surveillance system.



**Figure 17.56** A sticker or placard is a commonly used method of notifying the public that an area is under video surveillance.

The legislation or guidelines may cover the following:

- *License*: It may be necessary to register or obtain a license from an authority to conduct video surveillance, particularly in public areas.
- *Purpose of surveillance*: Is the equipment in accordance with what is permitted by the laws in the area?
- *Position or location*: Is the equipment located in such a way that it only monitors the spaces the equipment is intended to cover? If unintended areas are covered, consultations with the owners of such spaces might be required. In some areas, video surveillance can be prohibited, for example, restrooms and changing rooms in retail environments.
- *Notification*: Signs that warn the public that they are entering a zone covered by surveillance equipment might be necessary. Sometimes, the sign has to be of a particular type, or it may need to follow specific guidelines.
- *Quality of images*: There may be rules regarding the quality of images, which can affect what may be permitted or acceptable for use as evidence in court.
- *Video format*: Police authorities may require that the video format be one that they can handle.
- *Information provided in recorded video*: Video recordings may have to be stamped with time and date.
- *Processing of images*: There may be rules regulating how long images should be retained, who can view such images, and where recorded images can be viewed.
- *Drawings*: There may be requirements for drawings of where cameras are placed.
- *Personnel training*: There may be regulations that require operator training in security and disclosure policies as well as privacy issues.
- *Access to and disclosure of images to third parties*: There may be restrictions on who can access the images and how images can be shown. For example, if video will be disclosed to the media, images of individuals may have to be disguised or blurred.
- *Monitoring and recording of audio*: A permit may be required for recording audio in addition to video.
- *Regular system checks*: There may be guidelines on how often and thoroughly a company should perform system checks to make sure all equipment is operating as it should.
- *Audit trail*: Having the ability to show who used the system, at which time, and for what purpose. In addition, proof of a video's authenticity may be required by methods such as watermarking.





# Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>